

UNIVERSIDADE FEDERAL DE SANTA CATARINA

PROCESSAMENTO E REPRESENTAÇÃO DO
GÊNERO GRAMATICAL DO PORTUGUÊS:
Um mecanismo associativo

Natália C. A. de Resende

Florianópolis
Março de 2015

Ficha de identificação da obra elaborada pelo autor
através do Programa de Geração Automática da Biblioteca Universitária
da UFSC.

Resende, Natália Carolina Alencar de Resende
Processamento e Representação do Gênero Gramatical do
Português : um mecanismo associativo / Natália
Carolina Alencar de Resende Resende ; orientadora,
Mailce Borges Mota Mota ; coorientadora, Aline
Gesualdi Manhães. - Florianópolis, SC, 2015.

272 p.

Tese (doutorado) - Universidade Federal de Santa
Catarina, Centro de Comunicação e Expressão. Programa
de Pós- Graduação em Linguística.

Inclui referências

1. Linguística. 2. processamento do gênero
gramatical do português. 3. eletrofisiologia do
português. 4. simulação computacional do processamento
do gênero. 5. decisão de gênero. I. Mota, Mailce
Borges Mota. II. Gesualdi Manhães, Aline Gesualdi
Manhães. III. Universidade Federal de Santa Catarina.
Programa de Pós-Graduação em Linguística. IV. Título.

Natália Carolina Alencar de Resende

PROCESSAMENTO E REPRESENTAÇÃO DO
GÊNERO GRAMATICAL DO PORTUGUÊS DO BRASIL:
um mecanismo associativo

Tese apresentada como requisito final para obtenção do título de Doutor em Linguística pelo programa de pós-graduação em Linguística da Universidade Federal de Santa Catarina.

Orientador(a): Profa. Dra. Mailce Borges Mota.

Coorientador(a): Profa. Dra. Aline Gesualdi Manhães.

Florianópolis
Março de 2015

NATÁLIA CAROLINA ALENCAR DE
RESENDE

TÍTULO:

PROCESSAMENTO E REPRESENTAÇÃO DO
GÊNERO GRAMATICAL DO PORTUGUÊS:
Um mecanismo associativo

Esta tese foi julgada adequada para obtenção do
Título de “Doutor”, e aprovada em sua forma
final pelo programa de pós-graduação em
Linguística da Universidade Federal de Santa
Catarina.

Florianópolis, 26 de março de 2015.



Prof., Dr. Heronides Maurílio de Melo Moura,
Coordenador do Curso



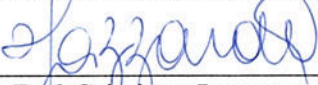
Prof.ª, Dr.ª Mailce Borges Mota,
Orientadora


Universidade Federal de Santa Catarina

Prof.ª, Dr.ª Aline Gesualdi Manhães.
Co-orientadora
CEFET-RJ


Banca Examinadora:


Prof.^a, Dr.^a Ana Cláudia de Souza,
Universidade Federal de Santa Catarina


Prof.^a, Dr.^a Cristiane Lazzaroto Volcão,
Universidade Federal de Santa Catarina


Prof.^a, Dr.^a Maria Luiza Gonçalves Aragão da
Cunha Lima,
Universidade Federal de Minas Gerais


Prof.^a, Dr.^a Maria Cristina Lobo Name,
Universidade Federal de Juiz de Fora


Prof.^a, Dr.^a Roberta Pires de Oliveira.
Universidade Federal de Santa Catarina

À vovó Isaura que não entende o que é um doutorado, mas sempre quis me ver ‘doutora’.

À memória de Simião da Costa Resende e Marilda de Alencar Travassos.

AGRADECIMENTOS

A caminhada foi longa, mas não foi solitária. Agradeço a todos os que direta ou indiretamente contribuíram para a realização deste trabalho e especialmente:

À professora Mailce Borges Mota por ter disponibilizado seu projeto de bolsa de produtividade em pesquisa junto ao CNPq, que serviu de base para a formulação das perguntas de pesquisa desta tese. Agradeço, também, o apoio recebido do LabLing, sob sua direção, pelos materiais concedidos para a realização das pesquisas, como o primeiro EEG usado no estudo eletrofisiológico conduzido para o trabalho de qualificação desta tese, o software E-Prime usado na apresentação dos estímulos, toda a infra-estrutura física do laboratório e, principalmente, pelo suporte técnico recebido dos bolsistas de iniciação científica do laboratório, os quais se mostraram sempre dispostos a contribuir. Não posso deixar de mencionar que sou grata, ainda, pelo incentivo da Profª. Mailce ao doutorado sanduíche cujo recurso financeiro foi fruto de um projeto paralelo entre o LabLing e o Max Planck Institute for Psycholinguistics em Nijmegen.

À professora Aline Gesualdi Manhães por ter conduzido os meus primeiros passos nos estudos eletrofisiológicos e por ter compartilhado sua *expertise* em *Machine Learning* ao contribuir com comentários valiosos e esclarecedores os quais foram fundamentais para a realização das simulações computacionais conexionistas apresentadas nesta tese.

Agradeço ao professor Pieter Seuren por ter, tão carinhosamente, acolhido minha chegada ao Max Planck Institute for Psycholinguistics. Agradeço ao professor Peter Hagoort por ter permitido meu acesso ao banco de dados de sujeitos da *Radboud University*, ao laboratório de EEG, às *research assistants* e a toda a infra-estrutura física do instituto. O meu mais sincero agradecimento a Daniel Acheson por ter me ensinado tudo o que sei hoje de EEG. Agradeço por ter se mostrado sempre entusiasmado, paciente como professor, e por ter sempre apresentado uma solução para os obstáculos e desafios que surgiram durante todo o processo. Agradeço imensamente pelas longas sessões de discussão dos dados, pelas dicas valiosas no manuseio do R e toda atenção concedida na orientação do experimento eletrofisiológico conduzido em Nijmegen. *Bedankt voor alles!*

Agradeço aos alunos de graduação e da pós-graduação da UFSC e aos alunos brasileiros de graduação e pós-graduação da *Radboud Universiteit* pela participação nas coletas de dados.

Agradeço aos amigos da Pós-graduação em Linguística e Pós-graduação em Inglês (em ordem alfabética), Anna, Gustavo, Joaquim, Karina, Luana, Mayra, Pâmela, Rafael, Sara e Viviane pelas demonstrações de amizade e companheirismo durante os 4 anos de doutorado. Ao lado de vocês, tudo foi mais divertido!

À Raquel Chaves pelo apoio intelectual, por ter sempre reservado um tempinho para me escutar e, principalmente, por ter contribuído para este trabalho com suas ideias construtivas e inspiradoras. Devo muito a você, amiga!!

Agradeço à minha avó materna, Isaura, pelas orações e palavras de incentivo. À minha irmã, Débora, por compreender minha ausência todos esses anos (mestrado e doutorado).

À Fernando pelo apoio e, principalmente, por não ter desistido de mim.

À Coordenação de Aperfeiçoamento de pessoal de nível superior (CAPES) pela concessão de apoio financeiro no Brasil e no exterior.

Agradeço a Deus por todas as bênçãos concedidas.

“I can’t go back to yesterday because I was a
different person then.”

Lewis Carroll, *Alice in Wonderland*

RESUMO

Um dos debates mais acalorados do campo da Psicolinguística e Neurocognição da Linguagem diz respeito ao processamento e representação mental de formas regulares e irregulares. De um lado, modelos de via dual (PINKER, 1991, 1999; ULLMAN, 2001a, 2001b, 2004, 2005) postulam que formas regulares, passíveis de decomposição, são processadas por um mecanismo responsável pela computação de regras. Formas irregulares, por sua vez, são armazenadas em sua forma pronta em um sistema de memória de caráter associativo. Por outro lado, modelos de via única (BYBEE, 1985, 1995) e modelos conexionistas (RUMELHART; McCLELLAND, 1986; PLUNKETT; JUOLA, 1999) preconizam que não há diferenças entre o processamento de itens lexicais regulares e irregulares. Ambas as categorias de forma são processadas por um único sistema de memória de natureza associativa. Esta tese teve como objetivo trazer contribuições empíricas para o debate em torno da dualidade vs. singularidade do processamento morfológico focalizando no processamento do gênero gramatical de substantivos inanimados do português do Brasil. Mais especificamente, investigamos se o processamento de formas regulares quanto ao gênero (substantivos femininos terminados em *-a*, e masculinos terminados em *-o*) distingue-se do processamento de formas irregulares (substantivos femininos terminados em *-o*, e masculinos terminados em *-a*), transparentes (substantivos com terminações típicas do gênero feminino e masculino como *-agem*, *-ade*, *-or*) e formas opacas quanto ao gênero (outras terminações). Para esse empreendimento, uma série de experimentos foi conduzida usando diferentes métodos de investigação, a saber: experimentos comportamentais, simulações conexionistas e um experimento eletrofisiológico. Os estudos comportamentais mostraram o papel crucial que a memória cumpre no processamento tanto da concordância quanto da atribuição de gênero a nomes substantivos inanimados independentemente de sua forma. Os resultados de três simulações computacionais usando um *perceptron* multicamadas mostraram que um sistema conexionista de via única é capaz de reproduzir certos padrões de processamento do gênero gramatical observados em falantes nativos do português. O estudo eletrofisiológico mostrou que o processamento da concordância de gênero com formas regulares, irregulares, opacas e transparentes recrutam componentes de ERP associados ao processamento automático da concordância morfossintática (LANs) e processos de re-análise e reparo da estrutura

sintática (P600s). Com base nos resultados obtidos, concluímos que o processamento do gênero gramatical de substantivos inanimados do português é governado por um único sistema neurocognitivo de natureza associativa.

Palavras-chave: gênero, português do Brasil, connexionismo, ERP, Psicolinguística, modelos de via dual, modelos de via única.

ABSTRACT

A hot debate within Psycholinguistics and Neurocognition of Language concerns the processing and mental representation of regular and irregular forms. On one hand, dual-route models (PINKER, 1991, 1999; ULLMAN, 2001a, 2001b, 2004, 2005) claim that regular decomposable forms are processed online by a mechanism responsible for the computation of rules. Irregular forms, however, are stored as whole words in an associative memory. On the other hand, single-route models (BYBEE, 1985, 1995) and connectionist models (RUMELHART; McCLELLAND, 1986; PLUNKETT; JUOLA, 1999) claim that there are no differences between the processing of regular and irregular forms. According to single-route models, all forms are processed by a single associative mechanism. This doctoral dissertation aims at bringing empirical contributions to the debate on the duality vs. singularity of morphological processing focusing on the processing of grammatical gender of inanimate nouns in Brazilian Portuguese. Specifically, we investigate if the gender processing of regular forms (feminine nouns ending in *-a*, and masculine nouns ending in *-o*) is distinct from the gender processing irregular (feminine nouns ending in *-o*, and masculine forms ending in *-a*), transparent (nouns whose endings are typical of feminine and masculine gender, such as: *-agem*, *-ade*, *-or*) and opaque forms (other endings). To this end, a set of experiments was conducted using different research methods, namely: behavioral experiments, connectionist simulations and an electrophysiological experiment. The results of the behavioral experiments revealed the role that memory plays in the processing of grammatical gender regardless of the noun's form. The results of three computational simulations using a multi-layer perceptron have shown that a single-route connectionist model can reproduce patterns of grammatical gender processing observed in native speakers of portuguese. The electrophysiological experiment indicates that the processing of gender agreement with regular, irregular, opaque and transparent forms recruits ERP components associated with automatic processing of morphosyntactic agreement (LANs) and with re-analysis and repair of the syntact structure (P600s). Based on these results, we conclude that the processing of grammatical gender of inanimate nouns in Brazilian Portuguese is governed by a single associative neurocognitive system.

Key-words: gender, Brazilian Portuguese, connectionism, ERP, Psycholinguistics, dual route models, single route models.

LISTA DE FIGURAS

FIGURA 1: APRESENTAÇÃO DOS ESTÍMULOS. EXPERIMENTO 1.....	92
FIGURA 2: <i>DIAGRAMAS DE CAIXA</i> DA VELOCIDADE DE RESPOSTA POR CATEGORIA DE ESTÍMULO. O (OPACAS), R (REGULARES).....	94
FIGURA 3: GRÁFICO DE INTERAÇÃO DE DOIS FATORES: FORMA E FREQUÊNCIA. EXPERIMENTO 1.....	96
FIGURA 4: APRESENTAÇÃO DOS ESTÍMULOS. EXPERIMENTO 2.....	103
FIGURA 5: <i>DIAGRAMAS DE CAIXA</i> DA VELOCIDADE DE RESPOSTA DE 30 SUJEITOS EM UMA TAREFA DE ATRIBUIÇÃO DE GÊNERO A 8 CATEGORIAS DE FORMAS: IRREGULARES FREQUENTES (IF) E INFREQUENTES (II); REGULARES FREQUENTES (RF) E INFREQUENTES (RI); TRANSPARENTES FREQUENTES (TF) E INFREQUENTES (TI); OPACAS FREQUENTES (OF) E INFREQUENTES (OI).	105
FIGURA 6: <i>DIAGRAMAS DE CAIXA</i> DA VELOCIDADE DE RESPOSTA DE 28 SUJEITOS EM UMA TAREFA DE ATRIBUIÇÃO DE GÊNERO A 48 PSEUDOPALAVRAS DISTRIBUÍDAS EM 2 GRUPOS: GRUPO DAS DIFÍCEIS À ESQUERDA E GRUPO DAS FÁCEIS À DIREITA.....	119
FIGURA 7: GRÁFICO DE INTERAÇÃO ENTRE O FATOR DIFICULDADE (FÁCEIS VS. DIFÍCEIS) E VELOCIDADE DE RESPOSTA POR TERMINAÇÃO DA PSEUDOPALAVRA.	121
FIGURA 8: GRÁFICO DE INTERAÇÃO ENTRE O FATOR DIFICULDADE (FÁCEIS VS. DIFÍCEIS) E VELOCIDADE DE RESPOSTA POR CATEGORIA DE PSEUDOPALAVRA.	121
FIGURA 9: : ORGANIZAÇÃO DO CONJUNTO DE DADOS: QUANTIDADE DE FORMAS/DISTRIBUIÇÃO DAS FORMAS EM INSTÂNCIAS (LINHAS).	139
FIGURA 10: COMPARAÇÃO DO DESEMPENHO DE QUATRO MODELOS COM DIFERENTES PARÂMETROS/COMPLEXIDADES.....	141
FIGURA 11: : CURVA ROC. SIMULAÇÃO 1.....	143
FIGURA 12: CURVA ROC. SIMULAÇÃO 2.....	146
FIGURA 13: CURVA ROC. SIMULAÇÃO 3.....	149
FIGURA 14: MONTAGEM DOS ELETRODOS INCLUÍDOS NA ESTATÍSTICA.....	176
FIGURA 15: MÉDIAS GLOBAIS DOS ERPs GERADOS PARA A CONCORDÂNCIA DE GÊNERO ENTRE DETERMINANTES E SUBSTANTIVOS (CONDIÇÃO 1). HEMISFÉRIO ESQUERDO. ONDAS NEGATIVAS PARA CIMA E POSITIVAS PARA BAIXO.....	178
FIGURA 16: MÉDIAS GLOBAIS DOS ERPs GERADOS PARA A CONCORDÂNCIA DE GÊNERO ENTRE DETERMINANTES E SUBSTANTIVOS (CONDIÇÃO 1).	

LINHA CENTRAL. ONDAS NEGATIVAS PARA CIMA E POSITIVAS PARA BAIXO.....	179
FIGURA 17: MÉDIAS GLOBAIS DOS ERPs GERADOS PARA A CONCORDÂNCIA DE GÊNERO ENTRE DETERMINANTES E SUBSTANTIVOS (CONDIÇÃO 1). HEMISFÉRIO DIREITO. ONDAS NEGATIVAS PARA CIMA E POSITIVAS PARA BAIXO.....	180
FIGURA 18: MÉDIAS GLOBAIS DOS ERPs GERADOS PARA A CONCORDÂNCIA DE GÊNERO ENTRE SUBSTANTIVOS E ADJETIVOS (CONDIÇÃO 2). HEMISFÉRIO ESQUERDO. ONDAS NEGATIVAS PARA CIMA E POSITIVAS PARA BAIXO.	181
FIGURA 19: MÉDIAS GLOBAIS DOS ERPs GERADOS PARA A CONCORDÂNCIA DE GÊNERO ENTRE SUBSTANTIVOS E ADJETIVOS (CONDIÇÃO 2). LINHA CENTRAL. ONDAS NEGATIVAS PARA CIMA E POSITIVAS PARA BAIXO....	182
FIGURA 20: MÉDIAS GLOBAIS DOS ERPs GERADOS PARA A CONCORDÂNCIA DE GÊNERO ENTRE SUBSTANTIVOS E ADJETIVOS (CONDIÇÃO 2). HEMISFÉRIO DIREITO. ONDAS NEGATIVAS PARA CIMA E POSITIVAS PARA BAIXO.....	183
FIGURA 21: MAPAS TOPOGRÁFICOS DO ESCALPO MOSTRANDO A DISTRIBUIÇÃO DOS ERPs NAS ÉPOCAS ANTERIORES (ENTRE 300MS E 500MS).FORAM ESCOLHIDAS JANELAS DE 50MS PARA CAPTURAR A DIFERENÇA DE PICO DE ERP ENTRE SENTENÇAS CONGRUENTES E INCONGRUENTES EM CADA CATEGORIA DE FORMA NA CONDIÇÃO 1.....	184
FIGURA 22: MAPAS TOPOGRÁFICOS DO ESCALPO MOSTRANDO A DISTRIBUIÇÃO DOS ERPs NAS ÉPOCAS ANTERIORES (ENTRE 300MS E 500MS). FORAM ESCOLHIDAS JANELAS DE 50MS PARA CAPTURAR A DIFERENÇA DE PICO DE ERP ENTRE SENTENÇAS CONGRUENTES E INCONGRUENTES EM CADA CATEGORIA DE FORMA NA CONDIÇÃO 2.....	185
FIGURA 23: MAPAS TOPOGRÁFICOS DO ESCALPO MOSTRANDO A DISTRIBUIÇÃO DOS ERPs NAS ÉPOCAS POSTERIORES (ENTRE 500MS E 900MS). FORAM ESCOLHIDAS JANELAS DE 50MS PARA CAPTURAR A DIFERENÇA DE PICO DE ERP ENTRE SENTENÇAS CONGRUENTES E INCONGRUENTES EM CADA CATEGORIA DE FORMA NA CONDIÇÃO 1.....	186
FIGURA 24: MAPAS TOPOGRÁFICOS DO ESCALPO MOSTRANDO A DISTRIBUIÇÃO DOS ERPs NAS ÉPOCAS POSTERIORES (ENTRE 500MS E 900MS). FORAM ESCOLHIDAS JANELAS DE 50MS PARA CAPTURAR A DIFERENÇA DE PICO DE ERP ENTRE SENTENÇAS CONGRUENTES E INCONGRUENTES EM CADA CATEGORIA DE FORMA NA CONDIÇÃO 2.....	187

LISTA DE TABELAS

TABELA 1: DESCRIÇÃO DA VELOCIDADE DE RESPOSTA POR CATEGORIA DE ESTÍMULO. EXPERIMENTO 1. DP (DESVIO PADRÃO).....	93
TABELA 2: VARIÁVEIS INCLUÍDAS NO MODELO QUE MELHOR EXPLICAM A VELOCIDADE DE RESPOSTA DE 19 SUJEITOS NO EXPERIMENTO 1.	95
TABELA 3: CATEGORIAS DE FORMAS E QUANTIDADE DE SUBSTANTIVOS POR TERMINAÇÃO NO EXPERIMENTO 2.	101
TABELA 4: DESCRIÇÃO DA VELOCIDADE DE RESPOSTA DE 30 SUJEITOS NA ATRIBUIÇÃO DE GÊNERO.....	104
TABELA 5: VARIÁVEIS INCLUÍDAS NO MODELO QUE MELHOR EXPLICAM A VELOCIDADE DE RESPOSTA DE 30 SUJEITOS NO EXPERIMENTO 2.	106
TABELA 6: CATEGORIAS DE FORMAS E QUANTIDADE DE PSEUDOPALAVRAS INCLUÍDAS NO EXPERIMENTO 3.	114
TABELA 7: PORCENTAGENS DE CLASSIFICAÇÃO DAS PSEUDOPALAVRAS EM CATEGORIAS DE GÊNERO: FÁCEIS X DIFÍCEIS.....	115
TABELA 8: DESCRIÇÃO DA VELOCIDADE DE RESPOSTA DE 28 SUJEITOS NA ATRIBUIÇÃO DE GÊNERO A PSEUDOPALAVRAS.....	117
TABELA 9: DESCRIÇÃO DA VELOCIDADE DE RESPOSTA DE 28 SUJEITOS POR TERMINAÇÃO: FÁCEIS VS. DIFÍCEIS. DP: DESVIO PADRÃO.....	118
TABELA 10: VARIÁVEIS INCLUÍDAS NO MODELO QUE MELHOR EXPLICAM A VELOCIDADE DE RESPOSTA DE 28 SUJEITOS NO EXPERIMENTO 3.	119
TABELA 11: REPRESENTAÇÃO NUMÉRICA DOS GRAFEMAS DOS.....	136
TABELA 12: CATEGORIAS DE FORMAS E QUANTIDADE DE SUBSTANTIVOS ESCOLHIDOS PARA CADA CATEGORIA.	137
TABELA 13: MATRIZ DE CONFUSÃO: CLASSIFICAÇÃO DA REDE VS. CLASSIFICAÇÃO REAL. TESTE PARA 603 INSTÂNCIAS.	142
TABELA 14: MATRIZ DE CONFUSÃO: CLASSIFICAÇÃO DA REDE VS. CLASSIFICAÇÃO DESEJADA. TESTE PARA 5 PSEUDOPALAVRAS.....	145
TABELA 15: MATRIZ DE CONFUSÃO: CLASSIFICAÇÃO DA REDE VS. CLASSIFICAÇÃO DESEJADA. TESTE PARA 6 PSEUDOPALAVRAS.	149
TABELA 16: CONTAGEM DE FREQUÊNCIA DE OCORRÊNCIA DOS BIGRAMAS SELECIONADOS (FREQUÊNCIA POR MILHÃO).....	168
TABELA 17: CONTAGEM DE FREQUÊNCIA DOS SUSBTANTIVOS SELECIONADOS (FREQUÊNCIA POR MILHÃO).	169
TABELA 18: QUANTIDADE DE ESTÍMULOS POR CATEGORIA E POR CONDIÇÃO.	170
TABELA 19: ELETRODOS INCLUÍDOS NA ESTATÍSTICA PARA CINCO REGIÕES DO ESCALPO	175

LISTA DE QUADROS

QUADRO 1: RELAÇÃO ENTRE ESTRUTURA FONOLÓGICA (TERMINAÇÃO) E GÊNERO MASCULINO SEGUNDO DESCRIÇÃO DE SAID ALI (1971)	49
QUADRO 2: RELAÇÃO ENTRE ESTRUTURA FONOLÓGICA (TERMINAÇÃO) E GÊNERO FEMININO SEGUNDO DESCRIÇÃO DE SAID ALI (1971).	50
QUADRO 3: CLASSIFICAÇÃO DOS NOMES EM PORTUGUÊS QUANTO À NATUREZA DO TRAÇO DE GÊNERO SEGUNDO NAME (2002).	59

Sumário

1 INTRODUÇÃO	23
1.1 ABORDAGEM E TERMOS USADOS NESTA TESE	28
1.2 OBJETIVOS	30
1.3 COMO ESTÁ ORGANIZADA ESTA TESE	31
2 A CATEGORIA GRAMATICAL DE GÊNERO DO PORTUGUÊS	39
2.1 O SUBSTANTIVO DE REFERENTE ANIMADO [± HUMANO]	39
2.1.1 FORMAÇÃO DO GÊNERO FEMININO DO PORTUGUÊS: FLEXÃO OU DERIVAÇÃO?	41
2.2 O SUBSTANTIVO DE REFERENTE INANIMADO [- HUMANO]	44
2.3 A CONCORDÂNCIA DE GÊNERO	54
2.3.1 ELEMENTOS MARCADOS PELO GÊNERO NO PORTUGUÊS	57
2.4 RESUMO DA DISCUSSÃO	59
3 MODELOS DE PROCESSAMENTO MORFOLÓGICO	63
3.1 INTRODUÇÃO	63
3.2 PINKER: <i>WORDS AND RULES</i>	64
3.3 ULLMAN: <i>MODELO DECLARATIVO/PROCEDURAL</i>	66
3.3.1 EVIDÊNCIAS: MODELOS DE VIA DUAL	69
3.4 BYBEE: <i>NETWORK MODEL</i>	72
3.4.1 EVIDÊNCIAS: MODELOS DE VIA ÚNICA	75
3.5 RESUMO DA DISCUSSÃO	77
4 PROCESSAMENTO DO GÊNERO GRAMATICAL DO PORTUGUÊS: COMPUTAÇÃO DE REGRAS OU ASSOCIAÇÃO DE PADRÕES ?	85
4.1 INTRODUÇÃO	85
4.2 EXPERIMENTO 1	90
4.2.1 METODOLOGIA	90
4.2.2 RESULTADOS	93
4.2.3 DISCUSSÃO DOS RESULTADOS	97
4.3 EXPERIMENTO 2	99
4.3.1 METODOLOGIA	100
4.3.2 RESULTADOS	104
4.3.3 DISCUSSÃO DOS RESULTADOS	107

4. 4 EXPERIMENTO 3	109
4.4.1 METODOLOGIA	113
4.4.2 RESULTADOS	115
4.4.3 DISCUSSÃO DOS RESULTADOS	122
4.5 DISCUSSÃO GERAL DO CAPÍTULO	126

5 CLASSIFICAÇÃO DE SUBSTANTIVOS DO PORTUGUÊS EM CATEGORIAS DE GÊNERO: UMA ABORDAGEM CONEXIONISTA

131

5.1 INTRODUÇÃO	131
5.2 A CONSTRUÇÃO DO CORPUS	135
5.3 O TREINAMENTO	139
5.4 SIMULAÇÃO 1	141
5.4.1 RESULTADOS DA SIMULAÇÃO 1	142
5.5 SIMULAÇÃO 2	144
5.5.1 RESULTADOS DA SIMULAÇÃO 2	145
5.6 SIMULAÇÃO 3	147
5.6.1 RESULTADOS SIMULAÇÃO 3	149
5.7 DISCUSSÃO GERAL DO CAPÍTULO	150

6 OS CORRELATOS NEUROFISIOLÓGICOS DO PROCESSAMENTO DA CONCORDÂNCIA DE GÊNERO DO PORTUGUÊS

157

6.1 INTRODUÇÃO	157
6.2 O PRESENTE ESTUDO	164
6.3 METODOLOGIA	167
6.4 PROCEDIMENTO DE COLETA DE DADOS DE ERP E ANÁLISE DOS DADOS	173
6.5 RESULTADOS	177
6.6 DISCUSSÃO DOS RESULTADOS	190

7 RESUMO E CONCLUSÕES

201

7.1 RESUMO	201
7.2 CONCLUSÕES	205

REFERÊNCIAS

209

APÊNDICE A: TERMO DE CONSENTIMENTO	223
APÊNDICE B: ESTÍMULOS EXPERIMENTO 1	227
APÊNDICE C: ESTÍMULOS EXPERIMENTO 2	230

APÊNDICE D: ESTÍMULOS EXPERIMENTO 3	235
APÊNDICE E: INPUTS APRESENTADOS À REDE NEURAL	238
APÊNDICE F: ESTÍMULOS EXPERIMENTO ELETROFISIOLÓGICO	241
APÊNDICE G: TERMO DE CONSENTIMENTO (EEG)	259
APÊNDICE H: DEBRIEFING	261
APÊNDICE I: PROTOCOLO DO EXPERIMENTADOR	262

CAPÍTULO 1

1 INTRODUÇÃO

“Gender is the most puzzling
of the grammatical categories”
Gender (Corbett, 1991)

A existência de um léxico mental responsável pelo armazenamento de pares arbitrários de som e significado é um consenso entre pesquisadores do campo da Psicolinguística e Neurocognição da Linguagem (PINKER, 1999; ULLMAN, 1999, 2001; BYBEE, 1985,1995, CHOMSKY, 1995; JACKENDOFF, 2002). Entretanto, a arquitetura e organização desse léxico tem sido, nas últimas décadas, um dos temas mais debatidos e pontos de controvérsias entre modelos de processamento da linguagem humana, conforme apontado por Bybee (1995), Alegre e Gordon (1999), Bayeen, Schreuder, De Jong e Krott (1999). No que se refere ao processamento dos aspectos morfológicos das línguas, dois modelos de processamento merecem destaque: os modelos teóricos de via dual (*dual-system theories*), ou dupla rota, e os modelos teóricos de via única (*single-system theories*), ou rota única.

Os modelos teóricos de via dual preconizam que a linguagem humana compreende dois sistemas cognitivos distintos: o léxico mental e a gramática mental (cf. PINKER 1991,1999; ULLMAN, 1999, 2001a). O léxico mental seria um grande repositório de signos linguísticos, isto é, seria composto por pares arbitrários de som (imagem acústica) e significado, em especial, por pares arbitrários de som e significado idiossincráticos e indecomponíveis em sua forma (por exemplo, verbos irregulares do inglês *go/went* ou do português *ser/era*). Esse sistema seria responsável, também, por associações entre itens da classe dos irregulares que apresentam regularidades fonológicas. Segundo Bybee (1995), regularidades fonológicas seriam traços compartilhados por pares de itens, como, por exemplo, as trocas de vogais para formar o passado do inglês (*sing-sang, ring-rang, fly-flew, blow-blew*). As formas irregulares que compartilham traços fonológicos são referenciadas na literatura como subregularidades fonológicas (cf. ULLMAN, 1999).

Em virtude do léxico mental ser um sistema de memória, itens frequentes armazenados no léxico mental seriam mais rapidamente acessados que itens infrequentes. Nesse sentido, formas presentes no

léxico mental estariam sujeitas a efeitos de frequência e processos de natureza associativa.

A gramática mental, por sua vez, seria um sistema com funções distintas do léxico mental. Para a visual dual de processamento, a gramática mental é um sistema responsável por operações de natureza computacional, tais como: regras sintáticas de combinação de itens em uma sentença e operações de natureza morfológica, como a concatenação de uma base a um afixo (*walk* + *ed* = *walked*), e regras morfossintáticas, como a concordância de gênero entre um substantivo e um adjetivo (*casa bonita*). De acordo com Pinker (1991, 1999) e Ullman (1999, 2001a, 2001b), apenas formas regulares, passíveis de decomposição, são processadas na gramática mental.

Os modelos de via única, em contrapartida, rejeitam a ideia de dissociação dos sistemas cognitivos para o processamento da linguagem. Para os pesquisadores desta vertente teórica, a linguagem humana compreende apenas um sistema de memória com funções e características similares ao léxico mental. Segundo os modelos de via única, todos os itens lexicais de uma língua, passíveis ou não de decomposição, seriam processados por uma rede integrada de conexões neurais. De forma semelhante, modelos conexionistas assumem que as formas simples e morfológicamente complexas estão representadas em um único sistema de memória e ligadas umas as outras por meio traços ortográficos, fonológicos e semânticos compartilhados (cf. BYBEE, 1995).

Em virtude dos pontos divergentes entre as duas perspectivas, ao longo das últimas décadas, diversos estudos surgiram com o objetivo de testar as hipóteses levantadas por um e outro arcabouço teórico. No entanto, grande parte desses estudos tem como foco de análise a representação e o processamento da flexão verbal das línguas naturais (por exemplo, BROVETTO; ULLMAN, 2005; ULLMAN, 1999) e, em menor proporção, a flexão de número (por exemplo, BAAYENS; DIJKSTRA; SCHREUDER, 1997; CLAHSSEN, EISENBEISS; SONNENSTUHL, 1997). Mais especificamente, esses estudos concentraram-se no processamento do contraste existente entre a flexão verbal e de número regular e irregular.

Nesta tese, pretendo dar prosseguimento a essa discussão. Proponho, como forma de testar as hipóteses dos modelos de via dual e modelos de via única, examinar a relação existente entre forma e gênero gramatical da variedade do português em uso no Brasil.

O estudo do gênero é particularmente interessante pelo fato dessa categorial gramatical ser, em várias línguas do mundo, incluindo o

português, um traço inerente a todo o conjunto de nomes, até mesmo inerente ao conjunto de nomes substantivos que denotam entidades inanimadas. Portanto, no português, e em várias línguas do mundo, assim como substantivos que designam seres sexuados como *médico*, *médica*, *gato*, *gata*, *menino*, *menina*, substantivos que têm como referente entidades não sexuadas, tais como: *dia*, *sol* e *razão*, *coragem* também estão distribuídos em categorias de gênero. Nesse sentido, a atribuição de gênero a nomes substantivos não está baseada apenas em uma propriedade semântica do referente, o seu sexo biológico, o gênero de nomes substantivos pode ser também um traço formal, historicamente fixado pelo uso (AZEREDO, 2013).

Há, por outro lado, línguas que adotam simplesmente critérios semânticos na atribuição de gênero aos nomes substantivos. O inglês é um exemplo. Em inglês, nomes substantivos que designam seres humanos do sexo masculino são do gênero masculino e, por isso, podem ser substituídos pelo pronome *he* (*man*, *son*, *nephew*, *uncle*); nomes substantivos que designam seres humanos do sexo feminino são nomes femininos e são substituídos pelo pronome *she* (*woman*, *daughter*, *niece*, *aunt*). Todos os outros nomes, inclusive os que designam seres não humanos, são neutros e são substituídos pelo pronome *it* (*cat*, *car*, *house*) (cf. CORBETT, 1991, p.12). Portanto, em inglês, o critério semântico para divisão dos nomes substantivos em categorias de gênero é o sexo biológico do referente humano.

Algumas línguas naturais, por sua vez, separam os nomes em categorias de gênero com base em outros critérios semânticos. Por exemplo, em Diyari, uma língua australiana aborígene, são masculinos todos os substantivos que denotam seres animados não humanos cuja distinção de sexo não é saliente (por exemplo, *peixe*), todos os substantivos que denotam entidades inanimadas, bem como todos os substantivos que denotam seres humanos do sexo masculino. Todos os substantivos que fazem referência a seres humanos do sexo feminino como *mulher*, *garota*, e substantivos que fazem referência a animais do sexo feminino com distinção de sexo saliente como *vaca* e *gata* são femininos. Assim, em Diyari, o critério semântico para divisão dos nomes em categorias de gênero inclui, também, o critério da saliência de sexo biológico do referente animado não humano (cf. CORBETT, 1991, p.11).

A partir dessas diferenças, Corbett (1991) separa as línguas do mundo em duas categorias em relação ao sistema de gênero: sistemas tipicamente semânticos e sistemas formais. Nos sistemas tipicamente semânticos, como vimos, além da correlação do gênero do substantivo

com o sexo biológico do referente, outros critérios são possíveis, tais como: os critérios humanos/não humanos, animados/não animados, sexo saliente/ sexo não saliente (CORBETT, 1991, p. 9). Entretanto, nas línguas de sistema formal, como as línguas românicas, a atribuição de gênero a um item lexical não pode ser baseada em critérios puramente semânticos visto que uma grande quantidade de nomes substantivos não se enquadra nesses critérios. Nessas línguas, os substantivos de referente inanimado podem estar distribuídos em duas (feminino e masculino) ou mais categorias de gênero (feminino, masculino e neutro). Muitas vezes, o gênero pode vir marcado na estrutura do substantivo, na maioria dos casos, em sua terminação. Por exemplo, no português, espanhol e italiano, como regra geral, formas terminadas em *-a* são, geralmente, femininas, enquanto as terminadas em *-o* são, geralmente, masculinas (cf. BATES *et al.* 1995; BARBER; SALLILAS, CARREIRAS, 2004).

Uma outra questão importante sobre o gênero é o fato de este traço se manifestar no comportamento de itens que estão em uma relação de associação (HOCKETT, 1958). Isso significa dizer que, no plano sintático, classes de palavras como artigos, adjetivos, numerais e pronomes devem concordar com o gênero do nome substantivo a que se referem (LUCCHESI, 2003, p. 430). Em um processo de concordância, adjetivos e determinantes podem receber uma indicação formal da informação de gênero em sua estrutura. No entanto, ao comparar os mecanismo de concordância de gênero entre as línguas do mundo, é possível observar que há uma variação entre os sistemas linguísticos. Embora seja comum que, em muitas línguas, adjetivos, artigos, numerais e pronomes concordem em gênero com o nome substantivo que acompanham, em outras línguas, verbos também podem concordar em gênero. Por exemplo, em Swahili e línguas Bantu em geral, o verbo concorda com o gênero do sujeito e pode concordar, também, com o gênero do objeto direto e indireto. (CORBETT, 1991, p.111).

É interessante notar que, além das diferenças existentes entre os sistemas de concordância, não só a forma dos substantivos varia entre as línguas, mas, também, seu gênero. Por exemplo, a palavra *mar* é masculina em português, em alemão é neutra (*das Meer*), e feminina em francês (*la mer*). Outro exemplo intrigante vem do hebraico. A forma para a palavra *lua* é *yare'ax*, uma palavra masculina, mas a palavra *levana*, que também significa *lua* em hebraico, é feminina (cf. GOLLAN, FROST, 2001). Essas diferenças revelam quão arbitrária é a distribuição de substantivos em categorias de gênero o que acaba por acarretar dificuldades no aprendizado desta categoria gramatical na segunda língua. No português como segunda língua, por exemplo, não é

raro observar construções agramaticais como **o noite*, **o saúde* **a sol* na fala ou mesmo produção escrita de aprendizes iniciantes e, até mesmo, de falantes em estágio mais avançado no idioma. É possível que os problemas enfrentados por aprendizes do português como segunda língua (principalmente por parte de aprendizes falantes nativos de línguas com sistemas de gênero puramente semânticos) na aquisição do gênero gramatical estejam relacionados a duas questões principais, conforme apontado por Taft e Meunier (1998) e Tucker, Lambert, Rigault e Segalowitz (1968): em primeiro lugar, como já mencionado, nem sempre é possível estabelecer uma correlação com a noção de sexo. Ao atribuir gênero ao substantivo *mulher*, os falantes recorrem a informação de sexo do referente, mas como saber se nomes como *saúde* ou *sol* são masculinos ou femininos se o significado dessas palavras não dizem nada a respeito de seu gênero?

Em segundo lugar, nem sempre é possível para o falantes do português como segunda língua estabelecer uma associação entre forma e gênero (cf. CORRÊA, 2001). Apesar de haver no português uma estreita relação entre a terminação *-a* e o gênero feminino, e a terminação *-o* e o gênero masculino, há uma quantidade considerável de formas que não se encaixam nesse padrão como *libido*, *foto* e *tribo* que são substantivos femininos, mas terminam em *-o*, e *poeta*, *cometa* e *mapa* que são substantivos masculinos, mas terminam em *-a*. Em acréscimo, certos substantivos do português não apresentam em sua estrutura fonológica ou ortográfica nenhuma pista para a dedução de seu gênero, tais como: *tarde*, *noite* e *leite*.

Em contraste, falantes nativos do português, ou de línguas que se enquadram no sistema formal de gênero, como francês, espanhol, italiano ou alemão, parecem cometer poucos ou nenhum erro de concordância (cf. TUCKER; LAMBERT; RIGAULT; SEGALOWITZ, 1968), sugerindo que, apesar da imprevisibilidade de um grande conjunto de formas em relação ao seu gênero, seu processo de aquisição parece ser tão eficaz quanto o processo de aquisição de outras categorias gramaticais (cf. HOOPER, 1980).

Tendo em vista as hipóteses de processamento levantadas pelos modelos de via dual e modelos de via única bem como as dificuldades enfrentadas na aquisição do gênero por parte dos falantes do português como segunda língua em oposição à rapidez e agilidade com que esta categoria gramatical é adquirida e processada na língua materna, as perguntas que se tornam relevantes são: Como os falantes nativos de uma língua como o português sabem e processam o gênero de uma quantidade considerável de substantivos de maneira tão ágil se a forma

ou significado de muitos desses substantivos não permite sua dedução? O processamento do gênero gramatical do português é melhor representado por um sistema dual ou por um único sistema de memória? Acredito que o estudo gênero pode trazer contribuições para o entendimento de como a informação linguística é armazenada e processada no léxico mental dos falantes do português como língua materna e, ainda, fornecer dados para um melhor entendimento de certos princípios organizacionais da cognição da linguagem humana.

1. 1 Abordagem e termos usados nesta tese

A terminologia adotada nesta tese está em consonância com os termos utilizados na literatura com foco na psicolinguística do gênero gramatical (por exemplo, CORBETT, 1991; TAFT; MEUNIER, 1998; GOLLAN; FROST, 2001; BARBER; CARREIRAS, 2003, 2005, BARBER; SALLILAS; CARREIRAS, 2004; HEIM, 2008; AFONSO *et al.*, 2013). Corbett (1991) ressalta que a atribuição de gênero aos nomes substantivos depende de dois tipos de informação: seu significado (informação semântica) e sua forma. A informação de gênero contida na forma, por sua vez, pode ser de dois tipos: informação contida na estrutura, como derivação e flexão de gênero, e informação contida na estrutura fonológica. Assim, segundo Corbett (1991, p. 33), a substantivos que apresentam a informação de gênero em sua estrutura morfológica, aplicam-se ¹regras morfológicas de atribuição de gênero, e a substantivos que apresentam a informação de gênero em sua estrutura fonológica, aplicam-se ²regras fonológicas de atribuição de gênero. Como o autor ressalta, nem sempre a diferença entre regras morfológicas e regras fonológicas é nítida, mas ainda é possível estabelecer uma distinção. Segundo Corbett (1991), pode-se dizer que, diferentemente das regras morfológicas, regras fonológicas de atribuição de gênero aplicam-se a apenas formas não derivadas, isto é, a substantivos primitivos. Exemplos de regras fonológicas de atribuição de gênero são: se um substantivo termina em vogal átona final *-a*, então é feminino (*casa, cadeira, borracha*), se um substantivo termina em vogal átona final *-o* (*carro, muro, brilho*), então é masculino. Exemplos de substantivos formados por regras morfológicas de atribuição de

¹do inglês: *morphological assignment rules*

²do inglês: *phonological assignment rules*

gênero são fornecidos por Azeredo (2013, p. 159). De acordo com Azeredo (2013), *claridade* e *escuridão* são substantivos femininos por força de uma regra que nos diz que são femininos todos os substantivos formados de adjetivos com acréscimo dos sufixos *-idade* e *-idão*, tais como: *felicidade*, *facilidade*, *aptidão*, *amplidão* (AZEREDO, 2013, p. 159).

Nesta tese, detenho-me na análise do processamento do gênero de substantivos que denotam entidades inanimadas. Os substantivos selecionados para os experimentos incluem formas derivadas, bem como substantivos primitivos, femininos ou masculinos por uma convenção, mas que podem (ou não) ser enquadrados em uma regra fonológica de atribuição de gênero em função de sua forma (estrutura fonológica).

É importante esclarecer que não é minha intenção discutir quais dos dois tipos de regras atuam sobre as formas. O meu objetivo aqui, como já afirmado, é investigar o impacto da estrutura dos substantivos (sua forma), principalmente no que diz respeito a sua terminação, no processamento (neuro)cognitivo do gênero do português. Os experimentos conduzidos para esta tese foram apresentados visualmente aos sujeitos participantes. O intuito foi ativar fonemas por meio da forma ortográfica dos substantivos apresentados visualmente para investigar o processamento de seu gênero. Nesse sentido, ao empregar o termo *Forma*, refiro-me ao que na literatura tem sido chamado de *estrutura fonológica* dos substantivos a qual, neste trabalho, foi ativada por meio de itens lexicais visualmente apresentados. Ao empregar o termo *forma (s) regular (es)*, refiro-me, portanto, a formas cuja estrutura fonológica ativada visualmente no léxico mental dos falantes pode ser submetida à regra (fonológica) de atribuição de gênero, e ao empregar *forma(s) irregular (es)*, refiro-me a formas cuja estrutura fonológica ativada visualmente contradiz a regra. Essa terminologia tem sido empregada por vários autores com esse mesmo sentido (cf. TAFT; MEUNIER, 1998, GOLLAN; FROST, 2001, AFONSO *et al.*, 2013, CORBETT, 1991, 2006).

Quanto às formas que não podem ser definidas como regulares ou irregulares, outros termos aparecem na literatura. Afonso *et al.* (2013) e Taft e Meunier (1998), por exemplo, fazem referência a substantivos *fonologicamente transparentes* àqueles cuja informação de gênero pode ser deduzida a partir de sua estrutura (em especial, sua terminação), e a substantivos *fonologicamente opacos* àqueles cuja informação de gênero não pode ser deduzida a partir de sua forma. Nesta tese, adoto também os termos empregados por Afonso *et al.* (2013) e Taft e Meunier (1998) – substantivos *fonologicamente*

transparentes e substantivos *fonologicamente opacos* – com o objetivo distinguir formas regulares e irregulares das formas que esses mesmos autores designam como fonologicamente transparentes e fonologicamente opacas quanto ao gênero. No capítulo 2, apresento, detalhadamente, as propriedades estruturais dessas categorias de formas (regulares, irregulares, transparente e opacas).

1.2 Objetivos

O objetivo geral desta tese é obter um melhor entendimento dos processos mentais que envolvem o gênero gramatical do português. O objetivo específico desta tese é investigar se o processamento do gênero gramatical do português é governado por um sistema dual ou um sistema unitário e, assim, contribuir com dados empíricos para o debate em torno do contraste existente entre o processamento de formas regulares, por um lado, e o processamento de formas irregulares, por outro. A minha hipótese inicial é a de que o gênero gramatical do português é governado por um mecanismo dual, isto é, formas regulares quanto ao gênero são processadas por um mecanismo de natureza computacional, ao passo que o gênero das demais formas é armazenado na memória.

Esta tese distingue-se de trabalhos anteriores que testaram as hipóteses levantadas pelos modelos teóricos de processamento morfológico da linguagem em diversos pontos. Em primeiro lugar, nossa língua de investigação é o português, em especial, a variedade brasileira do português, uma língua morfolologicamente rica que tem recebido relativamente pouca atenção por parte de pesquisadores da campo da Psicolinguística, principalmente no âmbito internacional. Em segundo lugar, apesar de grande parte dos estudos sobre o tema ter se concentrado na morfologia flexional de verbos conjugados no pretérito perfeito em língua inglesa, o presente estudo tem como foco o processamento do gênero gramatical de substantivos inanimados, um aspecto do português que também permite contrastar o processamento de formas regulares e irregulares (cf. Cap.2). Em terceiro lugar, esta tese lida com diferentes metodologias de pesquisa, a saber: métodos comportamentais, métodos de neuroimagem e simulações computacionais conexionistas. Empregar métodos e instrumentos de pesquisa diversos significou observar o processamento do gênero gramatical de substantivos inanimados do português a partir de

diferentes perspectivas a fim de buscar mais dados para uma interpretação mais precisa dos resultados.

Por fim, cabe ressaltar que esta tese contribui de maneira significativa para o programa de pesquisa coordenado pela Profa. Mailce Borges Mota, financiado com recursos de bolsa de produtividade em pesquisa, intitulado: *As assinaturas neurais da aquisição e processamento da morfologia flexional do PB: um estudo com falantes nativos e não nativos* (processo 307540/2010-3).

1.3 Como está organizada esta tese

Apresentadas a motivação, a terminologia e os objetivos desta tese, cabe expor sua organização. Esta tese está dividida em três partes: Parte I: Preliminares, Parte II: Experimentos e Parte III: Resumo e Conclusões. Na sequência desta introdução, na *Parte I*, apresento o enquadramento teórico-metodológico em que se situa esta tese. Essa parte preliminar compreende dois capítulos: Capítulo 2 e Capítulo 3. No Capítulo 2, *A categoria gramatical de gênero do português*, traço um percurso em que os pontos mais relevantes do sistema de gênero do português são destacados ao explorar seus aspectos morfológicos, fonológicos e sintáticos. O ponto importante desse capítulo é a descrição da relação existente entre a forma dos substantivos (de referente animado ou inanimado) e as categorias de gênero. Dessa descrição, veremos que certas formas podem ser agrupadas em diferentes categorias em virtude de compartilharem certos traços estruturais que as aproximam e, ao mesmo tempo, as diferem de outras formas. Agrupar as formas em categorias foi importante para a investigação do impacto que os traços compartilhados por elas exercem sobre o processamento do gênero no português. No capítulo 3, *Modelos de processamento morfológico*, apresento os modelos teóricos de interesse para o presente estudo: Modelos de via dual e Modelos de via única. Como já mencionado na *Introdução*, esses modelos apresentam hipóteses divergentes sobre o processamento da linguagem. As premissas desses modelos guiaram o desenho experimental dos experimentos apresentados na *Parte II* deste trabalho.

A *Parte II* é dedicada à apresentação dos experimentos que tiveram como objetivo contrastar o processamento das categorias de formas dos substantivos inanimados do português. Essa parte compreende três capítulos: Capítulo 4, Capítulo 5 e Capítulo 6. Em cada

um dos capítulos, reviso primeiramente os trabalhos relacionados ao experimento conduzido os quais adotaram métodos e procedimentos semelhantes. No Capítulo 4, apresento os dados de três experimentos comportamentais cujas análises foram baseadas no tempo de reação (velocidade de resposta) a estímulos linguísticos apresentados visualmente aos participantes. O objetivo geral desse capítulo é investigar se o processamento de formas regulares quanto ao gênero distingue-se do processamento de outras categorias de formas, a saber: transparentes, opacas e irregulares. Na sequência, Capítulo 5, apresento a construção e os resultados de um modelo conexionista que simula o processamento do gênero do português usando um *perceptron* multicamadas treinado com o algoritmo *backpropagation*. O Capítulo 6 tem como foco o estudo dos mecanismos neurocognitivos que são ativados durante o processamento da concordância de gênero do português. Para tanto, conduzi um experimento eletrofisiológico que contou com uma tarefa de leitura de sentenças apresentadas visualmente aos sujeitos participantes.

A terceira e última parte, *Parte III*, resume a pesquisa apresentada nos capítulos anteriores e conclui com uma discussão sobre as implicações dos achados desta tese para o debate em torno da singularidade vs. dualidade do processamento morfológico. Nesta parte, apresento também as limitações desta tese e sugestões de pesquisas futuras.

PARTE I: PRELIMINARES

PARTE I

PARTE I

Nesta primeira parte, apresento o enquadramento teórico-metodológico em que se situa esta tese. No Capítulo 2, apresento uma descrição do gênero do português, destacando sua manifestação tanto no plano sintático quanto na estrutura das formas. Como veremos a seguir, o gênero dos substantivos de referente animado diferencia-se do gênero dos substantivos de referente inanimado em alguns pontos. Por esse motivo, o gênero dos substantivos animados será descrito separadamente do gênero dos substantivos de referente inanimado para que contrastes possam ser estabelecidos. Considero essa primeira apresentação valiosa visto que os conceitos aqui descritos estão na base das questões por mim levantadas ao longo do trabalho. Em seguida, no Capítulo 3, ocupo-me dos principais construtos e hipóteses de representação e processamento dos modelos que constituem o referencial teórico no qual me apoio para delinear o desenho dos experimentos apresentados nos capítulos seguintes.

PARTE I

CAPÍTULO 2

2 A CATEGORIA GRAMATICAL DE GÊNERO DO PORTUGUÊS

2.1 O substantivo de referente animado [\pm humano]

No português, todos os nomes substantivos estão distribuídos em categorias de gênero: feminino ou masculino. A flexão de gênero, no entanto, restringe-se a um subconjunto de nomes substantivos que designam seres sexuados (ROCHA, 2008; CÂMARA, JR. 1970). Em nomes substantivos de referente sexuado, a terminação *-o* (*menin-o*, *vizinh-o*, *médic-o*, *gat-o*), e terminação *-a* (*menina-a*, *vizinh-a*, *médic-a*, *gat-a*) são morfemas flexionais de gênero (desinências de gênero) dado que são sufixos que expressam a oposição dos sexos feminino/masculino. Câmara Jr. (1970) ressalta que apenas formas femininas flexionam em gênero, já que as formas masculinas são formas não marcadas.

A oposição feminino/masculino no português pode acontecer, também, por heteronímia, ou seja, por meio de palavras de radical distinto, tais como: *homem/mulher*, *bode/cabra*. Nos pares *homem/mulher*, *boi/vaca*, *bode/cabra*, a oposição acontece somente ao nível da referência semântica, isto é, os falantes do português fazem uso apenas da oposição de significado entre o que é do sexo masculino (macho), e do sexo feminino (fêmea) para determinar o gênero do substantivo corretamente. Cabe ressaltar que o gênero desses vocábulos é invariável não se tratando, portanto, de flexão de gênero como em *gato/gata*; *médico/médica* (AZEREDO, 2013).

Uma outra forma de marcação do gênero na estrutura dos substantivos do português para oposição masculino/feminino se dá pelas desinências derivacionais, como nos exemplos a seguir:

- a) Diácono/Diaconisa - Maestro/Maestrina;
- b) Camponês/Camponesa-Freguês/Freguesa;
- c) Ermitão/Ermitoa;
- d) Cirurgião/Cirurgiã;
- e) Imperador/Imperatriz;
- f) Perdigão/Perdiz;
- g) Barão/Baronesa;
- h) Avô/Avó;
- i) Czar/Czarina.

Apesar de em muitos vocábulos do português a oposição feminino/masculino estar expressa na forma, seja por meio da desinência de gênero ou por desinências derivacionais, em vários outros casos, a oposição masculino/feminino fica a cargo dos determinantes (femininos e masculinos) que acompanham os nomes substantivos. Esses são os casos dos chamados substantivos *comuns de dois gêneros*: *o/a dentista*; *o/a herege*. Nesses casos, o gênero está expresso na estrutura sintática (CÂMARA JR., 1970).

Há, ainda, casos de substantivos que apresentam apenas um gênero para designar seres de ambos os sexos. Essa classe de substantivos é referenciada nas gramáticas dos português (cf. CUNHA; CINTRA, 2001; BECHARA, 2009) como a classe dos substantivos *sobrecomuns*, por exemplo, *o cônjuge*, *a testemunha*, *a vítima*. Nesses casos, o gênero do substantivo pode ser inferido do contexto semântico em que está inserido.

Os substantivos do português que designam seres animados não humanos em que a oposição de sexo não é saliente (epícenos) assumem apenas um gênero independentemente do sexo biológico do referente. Exemplos destes substantivos são: *a águia*, *a mosca*, *o peixe*, *o jacaré*. Nesse sentido, o gênero dos nomes *epícenos* assume um caráter formal, similar ao traço de gênero dos substantivos que designam entidades não animadas. Em ambos os casos, os substantivos apresentam um gênero intrínseco não variável: nomes como *a cobra*, *a águia* serão sempre femininos, e nomes como *o crocodilo*, *o corvo* serão sempre masculinos, independentemente de seu sexo biológico (CÂMARA JR., 1970, p. 89).

A outra similaridade entre os substantivos epícenos e os substantivos de referente não animado refere-se à relação entre terminação e gênero: assim como é possível observar uma relação entre terminação e gênero em substantivos de referente não sexuado (como será descrito a seguir), essa mesma relação está presente nos nomes epícenos. Por exemplo, tanto em *a mesa* e *o carro* quanto em *a cobra*, *a águia*, *o crocodilo*, a terminação é típica de seu gênero: terminação em *-a* (*cobra*, *mesa*, *águia*) para nomes femininos, e terminação em *-o* (*crocodilo*, *corvo*, *carro*) para nomes masculinos. Há, também, em ambos os casos, exceções. Como veremos na próxima seção, assim como há um subconjunto de nomes substantivos cuja terminação não é preditiva de seu gênero (*o leite*, *o abacaxi*), alguns substantivos *epícenos* também são opacos em sua forma quanto ao gênero (*o jacaré*, *o abutre*).

No português, a oposição feminino/masculino acontece, ainda, por alternância fonológica, como em *avô/avó*, e redução da forma do

masculino como em nomes terminados em ditongo, por exemplo: *órfão/órfã*; *anão/anã*; *irmão/irmã*; *campeão/campeã*, *réu/ré*.

Em resumo, a flexão de gênero no português, isto é, a oposição dos sexos feminino/masculino por meio de uma desinência de gênero é restrita a um subconjunto dos nomes substantivos de referente animado [\pm humano]. Em vários outros casos, a oposição de gênero se realiza ora por meio de radical diferente (por exemplo, *homem/mulher*), ora por meio de sufixos derivacionais (*barão/baronesa*), ora por meio de mudanças condicionadas morfofonologicamente, tais como: *réu/ré*, *avô/avó*. A oposição de gênero manifesta-se, ainda, somente no nível sintático por meio de determinantes e adjetivos variáveis, como: *a dentista talentosa*, *o dentista talentoso*.

Relativamente aos casos de flexão de gênero, cumpre esclarecer que há na literatura que trata da descrição do gênero do português uma discussão a respeito de seu status flexional ou derivacional. Abordo essa discussão na próxima seção.

2.1.1 Formação do gênero feminino do português: Flexão ou derivação?

Na seção anterior, mencionei que a flexão de gênero é privativa de um subconjunto de substantivos que designam seres animados. Contudo, uma das questões mais debatidas pelos gramáticos da língua portuguesa diz respeito à classificação do gênero como uma categoria gramatical de caráter flexional ou derivacional. Nessa discussão, a pergunta que se torna relevante é: qual a diferença entre flexão e derivação?

A flexão é definida por Azeredo (2013) como a variação da mesma forma. Segundo Rocha (2008) a flexão caracteriza-se por sua obrigatoriedade, sua sistematicidade e não-opcionalidade. No nível sintático, a flexão impõe normas de concordância. Por exemplo, os vocábulos associados a um substantivo devem com ele concordar em gênero e número. No caso dos verbos, o verbo deve, obrigatoriamente, concordar em número e pessoa com o sujeito de uma frase.

Quanto à sistematicidade, Rocha (2008) esclarece que, na língua portuguesa, as categorias gramaticais de pessoa, número, tempo e modo devem ser consideradas categorias flexionais pelo fato de grande parte dos verbos da língua seguirem um paradigma flexional regular. Em

relação às irregularidades, conforme aponta Rocha (2008), o conjunto de verbos irregulares é tão pequeno que pode ser considerado desprezível (ROCHA, 2008, p.195). Em relação à não-opcionalidade, o autor argumenta que um processo flexional não depende do falante criar novas formas. Por exemplo, para a flexão verbal, não depende do falante criar novas formas verbais de um determinado verbo. Caso um nova forma verbal entre na língua, o falante faz uso de morfemas flexionais pré-existentes.

Diferentemente da flexão, a derivação não é obrigatória. A derivação é um processo morfológico de criação de novas palavras a partir de um radical primário ao qual se acrescentam afixos e prefixos, como em: *re + ver = rever*; *feliz + idade = felicidade*. Há vários outros tipos de derivação, como derivação regressiva, derivação por sufixo zero e abreviação, as quais não serão discutidos aqui. No entanto, cabe aqui dizer que não é obrigatório que um falante faça uso de uma forma derivada para expressar uma ideia. Na frase ‘*eu vou reler os contratos amanhã*’, é possível que o falante use tanto a forma *reler* quanto *ler novamente* para expressar a mesma ideia.

Quanto à regularidade, os morfemas derivacionais do português são irregulares e assistemáticos (ROCHA, 2008). Como observa Câmara Jr. (1970, p.71), “uma derivação pode aparecer para um dado vocábulo e faltar para um vocábulo congênere”. Em relação à concordância, um processo derivacional não se manifesta no plano sintático, mas se estabelece entre uma nova palavra e outras palavras da língua. É uma relação paradigmática e não sintagmática (ROCHA, 2008, p. 205).

Em resumo, um processo flexional é caracterizado pela regularidade, obrigatoriedade da concordância e não-opcionalidade, ao passo que um processo derivacional é irregular e assistemático, não exige a concordância no plano sintático, e é opcional (cf. ROCHA, 2008). Embora seja possível estabelecer uma distinção entre processos flexionais e derivacionais, por que o debate permanece em relação ao gênero gramatical do português? Uma possível resposta para essa pergunta tem a ver com a dificuldade de se estabelecer se um dado vocábulo de gênero feminino como *menina* é uma forma flexionada de um mesmo vocábulo ou se é uma forma derivada de um forma primitiva. Como ressalta Rocha (2008, p.203), qual a diferença entre [[menino] S -*ada*] e [[galo] S -*inha*]?

Enquanto alguns gramáticos da língua portuguesa consideram que os nomes flexionam em gênero (CÂMARA JR., 1970; CUNHA; CINTRA, 2001), outros (AZEREDO, 2013; ROCHA, 2008) advogam a favor da tese da derivação. Azeredo (2013), por exemplo, assume que a

formação do feminino do português trata-se de um processo derivacional. O autor lista três razões para a defesa dessa tese:

- a) A quantidade de exceções à regra é incompatível com a ideia de regularidade imanente ao conceito de flexão;
- b) A variação das formas como *menino/menina* não são duas formas da mesma palavra, mas itens lexicais distintos e, por isso, apresentam entradas distintas nos dicionários do português;
- c) A criação de nomes femininos como *presidenta*, *mestra*, não são processos obrigatórios, mas opções estilísticas dos falantes, o que não acontece com processos de flexão regular.

Diferentemente de Azeredo (2013), Villalva (2000) não apresenta uma posição explícita a favor da ideia do gênero feminino do português ser formado por um processo derivacional. Entretanto, a autora defende que a realização dos contrastes de gênero não pode ser atribuída a processos flexionais, independentemente de o nome se referir ou não a seres sexuados. Villalva (2000) assume essa posição com base nas seguintes razões:

- a) Os contrastes de gênero não afetam a totalidade dos nomes substantivos;
- b) Os contrastes de gênero não são obrigatórios e a sua realização não é sistemática;
- c) Os sufixos flexionais não variam, ao passo que os sufixos indicadores de gênero são vários.

Villalva (2000) destaca que, apesar de os contrastes de gênero não se enquadrarem na categoria flexional, o gênero é uma propriedade inerente dos nomes substantivos. Por esse motivo, a autora argumenta que os contrastes de gênero são lexicalmente determinados, ou seja, especificados no léxico mental dos falantes, visto que não sofrem processos de flexão. Segundo a autora, a morfologia de gênero do português é um processo puramente morfossintático (cf. VILLALVA, 2000, p. 234).

Quanto aos processos de formação do gênero por derivação, Villalva (2000, p.241) argumenta que há uma redundância na descrição do gênero com referência à derivação: na maioria dos casos de gênero formado por sufixação derivacional, o feminino apresenta tema em *-a* (*camponês/camponesa*, *maestro/maestrina*). Quanto aos demais, mesmo

que os masculinos sejam atemáticos (*peru/perua*) ou tema zero (*juiz/juíza*), o feminino é formado com tema em *-a*.

Rocha (2008), por sua vez, assume que a formação do gênero feminino do português trata-se de um processo derivacional, por um lado, e de um processo flexional por outro lado. Para Rocha, há flexão de gênero no plano sintático. Nesse sentido, o autor deixa de lado a posição de flexão adotada por Câmara Jr. (1970) e Azeredo (2013), que define como flexão a variação de um mesmo vocábulo, e adota a posição de Anderson (1992, p. 74), que define como flexão a interação de regras sintáticas e morfológicas. Assim, para Rocha (2008), no plano sintático, os nomes substantivos flexionam em gênero em virtude de haver concordância de gênero obrigatória entre um substantivo e um adjetivo e entre um substantivo e um determinante. No que se refere à concordância, Rocha (2008) assume que não há diferenças entre os processos em operação no pares *menino/menina* e *menino/meninos*. Para o autor, em ambos os exemplos, há um processo derivacional operando. No entanto, no caso do número (*menino/meninos*), a regra de derivação é regular e não apresenta restrições, mas, no caso do gênero, é bastante irregular.

Em suma, na literatura que trata da descrição do gênero do português é possível encontrar um debate sobre a classificação da categoria de gênero como um processo flexional ou derivacional. Enquanto, tradicionalmente, a formação do feminino é tratado como um processo flexional, alguns autores defendem que a realização dos contrastes de gênero trata-se de um processo derivacional, enquanto outros defendem que há, por um lado, derivação e, por outro, flexão.

2.2 O substantivo de referente inanimado [- humano]

Como vimos, o gênero dos substantivos que designam seres sexuados pode estar ou não morfológicamente marcado por meio de uma desinência de gênero. De fato, apenas um subconjunto dos nomes substantivos sexuados apresentam marca de gênero em sua estrutura. No entanto, ao atribuir gênero aos substantivos que tem como referente um ser sexuado, os falantes do português têm como referência uma propriedade semântica: o seu sexo biológico. Essa propriedade semântica é o fator desambiguador na ausência de pistas estruturais (morfemas, desinências) de gênero. Por outro lado, e por razões óbvias, o fator de ordem semântica não se aplica aos nomes substantivos de

referente não sexuado, seja este concreto ou abstrato, como: *casa, cultura, carro, volume, cidade, canção*. Como já mencionado, em português, todos os nomes substantivos têm um gênero imanente (feminino ou masculino), independentemente de fazer referência a um ser sexuado ou não.

Rocha (2008) mostra que os substantivos não sexuados são a maioria: 95% dos substantivos do português designam seres não sexuados, ao passo que apenas 4,5% designam seres sexuados, incluindo aqueles que não apresentam uma marca morfológica para atribuição de gênero (*criança, homem, selvagem*, etc). Assim sendo, a pergunta que se faz é: Como os falantes nativos do português sabem o gênero de tantos substantivos se esse traço não apresenta nenhuma correlação com a noção de sexo?

A resposta para essa pergunta está longe de ser trivial, mas é preciso mencionar que um grande conjunto de vocábulos apresenta pistas de seu gênero em sua estrutura (morfo)fonológica. Nesse sentido, o critério para a atribuição de gênero aos nomes deixa de ser semântico. Critério formais de atribuição de gênero aos nomes entram em cena (cf. CORBETT, 1991). Por exemplo, assim como as desinências de gênero -a e -o formam substantivos femininos e masculinos, uma grande parte dos substantivos do português que designam seres não sexuados femininos termina em vogal temática (átona final) -a, e os que designam seres masculinos terminam em vogal temática (átona final) -o. Como observa Villalva:

Esta redistribuição dos nomes pelas classes temáticas tende a privilegiar uma correlação com o gênero, estabelecendo um vínculo preferencial entre tema em -a / gênero feminino (cf. *luxúria, pôpa, grua*) e tema em -o / gênero masculino (cf. *dono, sogro, genro, livro, pássaro, corpo, tempo*) (VILLALVA, 2000, p. 236)

É importante ressaltar que, diferentemente dos substantivos que designam seres animados, as terminações -a e -o desses vocábulos não são morfemas, mas vogais temáticas (VILLALVA, 2000; MARGOTTI, 2008), visto que essas não cumprem o papel de distinção de sexo como cumprem em vocábulos do tipo: *menino/menina, gato/ gato*. A totalidade dos nomes substantivos dessa classe pertence a um gênero único (cf. ROCHA, 2008). Assim, *dedo* é um substantivo de gênero somente masculino. Esse substantivo não apresenta uma contraparte do

sexo feminino como *deda**. Da mesma forma, *cultura* é um substantivo do gênero unicamente feminino. Portanto, esses substantivos não flexionam gênero e, por isso, não é possível falar, nesses casos, em regras de flexão (ou derivação, segundo alguns autores)³ morfológica. Como ressalta Corbett (1991, p. 51), quando o gênero em questão é específico de uma única forma, então estamos lidando com uma regra fonológica de atribuição de gênero e não uma regra morfológica. Sendo assim, segundo o autor, a regra da terminação em *-a* para nomes substantivos femininos, e *-o* para nomes substantivos masculinos que têm como referente uma entidade não sexuada é uma regra fonológica de atribuição de gênero. Para Corbett (1991), uma regra fonológica seria, por exemplo: se o substantivo termina em vogal átona final *-a*, então o substantivo é feminino; se o substantivo termina em vogal átona final *-o*, então o substantivo é masculino.

Por outro lado, conforme já mencionado anteriormente, o gênero de palavra derivadas formadas em um processo de sufixação pode ser resultado de uma regra morfológica. Nos pares *feliz/felicidade*; *pontual/pontualidade*; *fácil/facilidade*, é possível observar que a adição do sufixo *-idade* a adjetivos como *feliz*, *pontual*, *fácil* forma substantivos do gênero feminino. O sufixo *-agem*, formador de substantivos abstratos (ROCHA, 2008, p.107), também forma substantivos femininos. Embora o gênero desses substantivos seja determinado por força de uma regra morfológica (AZEREDO, 2013), é possível encaixar a atribuição de gênero a esses nomes em uma regra fonológica como por exemplo: nomes substantivos terminados na sequência de fonemas (representados pelos grafemas) *-idade* e *-agem* são femininos. Conforme observa Corbett (1991), nem sempre a distinção entre regras morfológicas e fonológicas de atribuição de gênero é clara. Nas palavras do autor:

[...] um grande número de substantivos não se enquadra nas regras semânticas de atribuição. No entanto, esses substantivos podem ser enquadrados em regras formais de atribuição, isto é, regras que dependem mais da forma dos nomes envolvidos do que de seu significado. Essas regras são de dois tipos: morfológicas e fonológicas. Embora a distinção entre regras semânticas e formais de atribuição de gênero seja clara (apesar de uma sobreposição de seus efeitos

³ Por exemplo: Azeredo (2013)

ser possível), a distinção entre regras morfológicas e fonológicas não é sempre evidente. Como regra geral, pode-se dizer que regras fonológicas de atribuição referem-se apenas à forma única de um nome, por exemplo, ‘nomes terminados em uma vogal no singular são femininos’. Tipicamente, a forma mais básica do nome está envolvida, apesar de nem sempre ser pronunciada. Regras morfológicas, por outro lado, exigem mais informação; eles referem-se a mais de uma forma. Isto nem sempre é óbvio.⁴(CORBETT, 1991, p. 33) (tradução nossa).

De fato, é preciso considerar que a distinção entre regras morfológicas e fonológicas de atribuição de gênero torna-se complexa na medida em que estamos lidando com uma interface existente entre ambos os aspectos da língua. Como visto, embora o gênero de muitos vocábulos do português seja produto de uma regra morfológica, como a adição de sufixo à base de um adjetivo transformando-o em substantivo e, ao mesmo tempo, distribuindo-o em uma categoria de gênero, estamos lidando, também, com uma regra fonológica de atribuição de gênero. Trata-se, de fato, de uma mudança morfofonêmica. Conforme explica Zanotto (1986, p. 40-41), uma mudança morfofonêmica se dá quando há uma mudança no sistema fonêmico do vocábulo, com repercussão no sistema mórfico.

No que concerne a relação entre estrutura fonológica dos substantivos e gênero, Said Ali (1971) apresenta uma descrição do gênero do português de caráter diacrônico, levando em consideração a terminação, isto é, os fonemas finais dos vocábulos e a relação desses fonemas finais com o gênero que os vocábulos assumem. Em um

⁴ ‘[...] large numbers of nouns fall outside the semantic assignment rules. These nouns may be handled instead by formal assignment rules, that is, rules which depend on the form of the nouns involved rather than on their meaning. These rules are of two types, morphological and phonological. Whereas the distinction between semantic and formal assignment rules is clear (though their effects may overlap), the distinction between morphological and phonological rules is not always clear-cut. As a rule of thumb, we may say that phonological rules refer just to a single form of a noun, for example, ‘nouns ending in a vowel in the singular are feminine’. Typically, the most basic form of the noun is involved, though this is not always spelled out. Morphological rules, on the other hand, require more information; they need to refer to more than one form. This is not always obvious.

primeiro momento, o autor esclarece, assim como Cunha e Cintra (2001), que os vocábulos tipicamente femininos em português são os terminados em vogal *-a* átona final, e os masculinos são os terminados em vogal *-o* átona final. Em relação aos processos atípicos de formação do gênero, o autor apresenta uma análise contrastando a formação do feminino ao longo dos séculos em português. Said Ali (1971) descreve as terminações e mostra as transformações que sofreram do latim. Por exemplo, os termos de entidades inanimadas abstratas terminados em *-io* no latim transformaram-se em femininos no português com a terminação em *-iã* (*legião, opinião*), assim como termos em *-zão, -ção, -são* que terminavam em *-tio* e *-sio* no latim (*condição, razão, fusão, produção*). Em acréscimo, o autor lista vários vocábulos que eram femininos e se transformaram em masculinos (*mapa, fantasma*), masculinos que são hoje femininos (*catástrofe, pirâmides*) e, ainda, os de gênero incerto no século XVI (*planeta, cometa, tribo*), além dos que mantiveram o gênero como os terminados em *-agem* (*linguagem*) que são sempre femininos.

O ponto importante da exposição de Said Ali (1971) é a relação de caráter preditivo que o autor estabelece entre as terminações dos vocábulos e o gênero que assumem na língua. Os Quadros 1 e 2, a seguir, ilustram a sistematização das terminações dos substantivos e sua relação com a atribuição de gênero segundo Said Ali (1971).

Quadro 1: Relação entre estrutura fonológica (terminação) e gênero masculino segundo descrição de Said Ali (1971)

Gênero	substantivos terminados em:	Exemplos	Exceções
Masculino	<i>-ão</i>	Chão, grão, sabão (concretos)	Televisão
	<i>-ão</i>	carrão, portão,mulherão (aumentativos)	
	<i>-ote</i>	filhote, mascote, pacote	glote, epiglote
	<i>-nte</i>	monte, horizonte	fonte, ponte, fronte
	<i>-arte</i>	estandarte	martre, parte,arte
	oxítonos	pé, anzol,filó,clã	pá,fé,maré,ralé ,libré, enxó,cor, dor, flor, lã,romã,avelã, hortelã,nau,lei, cicatriz,matriz, raiz,foz, noz,tardoz,voz cruz,luz,paz,te naz,aguarráz

Quadro 2: Relação entre estrutura fonológica (terminação) e gênero feminino segundo descrição de Said Ali (1971).

Gênero	Terminação	Exemplo	Exceção
Feminino	<i>-agem</i>	imagem, viagem, ferrugem	
	<i>-ão</i>	legião, ocasião, opinião (abstratos)	mão (concreto)
	<i>-ção</i>	reação, civilização, bênção (abstratos)	
	<i>-são</i>	fusão, missão, visão (abstratos)	televisão (concreto)
	<i>-zão</i>	razão, coalizão, vazão (abstratos)	
	<i>-ade</i>	cidade, variedade, mocidade	
	<i>-ice</i>	velhice, chatice, burrice	
	<i>-ez</i>	altivez, solidez, vez	
	<i>-ude</i>	saúde, altitude, fraude	
	<i>-al</i>	catedral, pastoral, diagonal	
	<i>-ede</i>	sede, rede, parede	

Gênero	Terminação	Exemplo	Exceção
	<i>-ide</i>	vide, lide,pirâmide,lápide	
	<i>-ave</i>	chave, trave, clave,	conclave
	<i>-eve</i>	neve, greve	
	<i>-ebe</i>	plebe, sebe	
	<i>-ase</i>	base, gase, fase	
	<i>-ace</i>	face, alface	
	<i>-ece</i>	prece	
	<i>-esse</i>	quermesse, messe	interesse, desinteresse, estresse
	<i>-ese</i>	tese, diocese,catequese, síntese	
	<i>-ose</i>	apoteose,dose,osmos e	
	<i>-ente</i>	enchente, torrente, corrente	
	<i>-pole</i>	Metrópole, Necrópole	
	<i>-ite</i>	Bronquite, Fulgurite, Dematite	

Gênero	Terminação	Exemplo	Exceção
	-orte	Sorte, Morte, Corte	

Como se pode observar no Quadro 1 e Quadro 2, Said Ali (1971) descreve a relação entre terminação dos nomes substantivos do português e seu gênero gramatical. Entretanto, apesar de Said Ali enfatizar a relação existente entre várias terminações e sua relação com o gênero que assumem, algumas terminações típicas do gênero masculino não são mencionadas pelo autor. Por exemplo, como apontam Cunha e Cintra (2001), são masculinos os substantivos de origem grega terminados em *-ema* e *-oma* (*fonema, problema, cinema, idioma, sintoma, aroma*). Além dessas, é possível estabelecer também uma relação entre o gênero masculino e outras as terminações, tais como: *-or* (*setor, calor, vapor, terror*) e *-ume* (*perfume, costume, ciúme, legume*). Em uma busca no *Corpus Brasileiro*⁵, apenas os vocábulos *flor, dor* e *cor* foram encontrados como exceção à relação entre a terminação *-or* e o gênero masculino. Em relação à terminação *-ume*, apesar de infrequente (apenas⁶ dez vocábulos contendo a terminação *-ume* foram encontrados) todos os substantivos terminados em *-ume* pertenciam ao gênero masculino.

A relação das terminações acima elencadas e as categorias de gênero revela que, além da regra geral (substantivos terminados em *a* átono são femininos, e substantivos terminados em *-o* átono são masculinos), é possível encontrar no português outros padrões estruturais que apresentam uma relação de transparência com as categorias de gênero, como: *-ema, -ume, -agem, -idade/ade*. Contudo, a relação dessas terminações com os gêneros feminino e masculino distingue-se da relação da regra geral em pontos importantes:

a) As terminações típicas do gênero feminino e masculino acima elencadas (cf. Quadro 1 e Quadro 2) são menos frequentes que as

⁵ O *Corpus Brasileiro* representa uma coletânea de textos escritos em português do Brasil e contém, aproximadamente, um bilhão de palavras e vários gêneros textuais. A descrição do corpus e dados estatísticos estão disponíveis no site da *Linguateca*: www.linguateca.pt.

⁶ *volume, legume, costume, lume, perfume, curtume, vagalume, cume, tapume, azedume*

terminações *-a* para o gênero feminino e *-o* para o gênero masculino. Por exemplo, em uma busca no *Corpus Brasileiro*, uma coletânea de textos contendo mais um bilhão de palavras, apenas 872 valores de lema de substantivos com terminação em *-agem* foram encontrados. Para substantivos terminados em *-ade* foram encontrados 4363 valores de lema, substantivos terminados em *-ume* e *-ema*, terminações típicas do gênero masculino, foram encontrados somente 256 e 832 valores de lema respectivamente;

b) No plano sintático, *-a* e *-o* são as terminações que exprimem a relação de concordância de gênero entre substantivos e adjetivos variáveis, ao passo que as outras terminações não exercem essa função;

c) São terminações variadas e distintas para uma mesma categoria de gênero e estão presentes em um conjunto pequeno de nomes, ao passo que *-a* e *-o* estão presentes em um conjunto extenso de nomes substantivos inanimados. Em acréscimo, conforme mostra Nascimento (2006), *-a* e *-o* são as terminações *default* para a atribuição de gênero a não palavras, ao passo que as terminações apresentadas nos Quadros 2 e 3 não são tão produtivas para a formação do gênero no português;

d) São transparentes quanto ao gênero. Em contrapartida, a transparência das terminações *-a* e *-o* é instável, pois, como já mencionado, apesar da estreita relação existente entre terminação *-a* e o gênero feminino, e a terminação *-o* e o gênero masculino, há vocábulos no português que contradizem essa regra como *a libido*, *a mão*, *o cometa*, *o fantasma*.

Com base nos traços estruturais compartilhados pelos substantivos inanimados do português e a relação desses traços com as categoria de gênero (feminino/masculino), é possível agrupar os substantivos do português em categorias de forma quanto ao seu gênero gramatical. As categorias são:

Regulares: formas femininas terminadas em vogal átona final *-a*, e masculina terminadas em vogal átona final *-o*: *casa*, *vida*, *água*, *força*, *livro*, *futuro*, *carro*. Essas formas são consideradas na literatura como regularidades fonológicas quanto ao gênero (cf. CORBETT, 1991; AFONSO *et al.*, 2013);

Irregulares: formas cuja terminação contradiz a expectativa de gênero. No caso do português, formas femininas terminadas em vogal átona final *-o*, e masculina terminadas em vogal átona final *-a*. Essas formas são consideradas na literatura como irregularidades fonológicas quanto ao gênero (cf. TAFT; MEUNIER, 1998; AFONSO *et al.* 2013). É possível acrescentar nessa categoria formas com terminação em *-ão*.

Apesar de ser uma terminação tônica, no português, um grande conjunto de substantivos inanimados terminam em *-ão* e são femininos. No entanto, muitos substantivos masculinos também terminam em *-ão*. Assim sendo, substantivos terminados em *-ão* podem ser considerados irregulares na medida em que essa terminação é ambígua, isto é, pode estar associada tanto ao gênero masculino quanto ao gênero feminino.

Transparentes: terminações típicas do gênero feminino como *-agem* e *-idade* (*garagem, cidade*), e do gênero masculino, como *-or*, *-ume*, (*perfume, setor*) (cf. AFONSO *et al.* 2013).

Opacas: formas cuja terminação não apresentam pistas para dedução de seu gênero: *abacaxi, leite, tarde, noite* (cf. AFONSO *et al.*, 2013).

Vale lembrar que, como apontado na introdução, a terminologia adotada nesta tese está em consonância com os termos encontrados na literatura que tem como foco o processamento do gênero gramatical em outras línguas. Apesar de autores, como Name (2002), adotarem o termo *(ir)regularidades fônicas* para fazer referência às relações que se pode estabelecer entre terminação e gênero, adoto aqui os termos *regularidades fonológicas, irregularidades fonológicas, transparência fonológica* e *opacidade fonológica* em relação ao gênero, buscando manter a terminologia adotada em estudos empíricos que investigam o processamento do gênero gramatical para outras línguas os quais adotaram métodos e procedimentos de pesquisa semelhantes ao presente trabalho (cf. Cap.4, Cap. 5 e Cap. 6).

2.3 A concordância de gênero

Nas seções anteriores (seção 2.1 e seção 2.2), vimos que o todos os substantivos do português apresentam um gênero imanente: feminino ou masculino. A atribuição de gênero aos nomes pode ser baseado em um critério semântico, o sexo biológico do referente, e pode ter como base, também, (e ao mesmo tempo) critérios formais, como regras morfológicas (adição de sufixos a uma base que distribuem os substantivos em categorias de gênero) ou regras fonológicas de atribuição de gênero (vogais terminais dos substantivos). Nesta seção, volto minha atenção para a descrição da concordância de gênero do português. Essa descrição é importante pelo fato de ser por meio da concordância que o gênero se realiza na língua portuguesa.

No plano sintático, o gênero dos substantivos do português, seja este determinado por um critério semântico ou um critério formal,

especifica o gênero dos determinantes (artigos, pronomes, numerais) e adjetivos, isto é, determinantes e adjetivos devem concordar em gênero e número com o substantivos a que se referem. No entanto, quando o substantivo designa um ser sexuado que permite estabelecer a oposição masculino/feminino por meio de uma desinência de gênero (*gato/gata, menino/menina*), a regra morfológica de flexão (ou derivação) de gênero é acionada. Assim, a concordância de gênero fica condicionada ao gênero que o substantivo assume. Por exemplo, temos *o gato esperto*, mas é possível ter, também, *a gata esperta*. Nesses exemplos, *gato* e *gata* assumem gêneros diferentes, visto que indicam a oposição de sexo masculino/feminino. Essa oposição está marcada na estrutura morfológica do vocábulo e tem repercussões no plano sintático. Como o substantivo *gato* em '*o gato esperto*' assumiu o gênero masculino, então os determinantes e adjetivos receberão a informação de gênero masculino. Essa informação vem marcada nas estrutura dos adjetivos (*espert-o/espert-a*) e nos determinantes (cf. ROCHA, 2008; AZEREDO, 2013).

No que se refere a substantivos de referente não sexuado ou referente sexuado cujo diferenciação de sexo não é saliente (epícenos), a marcação do gênero do substantivo se dá estritamente na concordância. Por exemplo, na frase '*O jacaré é perigoso*' o gênero (masculino) do substantivo determina a seleção da forma masculina do determinante e acrescenta a desinência de gênero masculina no adjetivo. O mesmo comportamento se observa em relação aos substantivos de referente não sexuado. Na frase '*A pequena ponte da cidade*' o gênero do substantivo *ponte* (feminino) é imanente, convencionado pelo uso e determina a seleção da forma feminina do determinante e acrescenta a desinência de gênero feminina no adjetivo *pequena*. A partir das formas dos substantivos *ponte* e *jacaré*, o seu gênero não pode ser deduzido, mas está expresso no nível sintático, por meio da concordância (ROCHA, 2008).

Por outro lado, um grande subconjunto de substantivos femininos (final *-a*) e masculinos (final *-o*) de terminação regular quanto ao gênero, bem como o subconjunto de substantivos com terminação típica do gênero feminino (por exemplo, final *-agem, -idade*) e masculino (por exemplo, final *-or, -ume*) apresentam em sua estrutura fonológica pistas para a dedução de seu gênero. Desse modo, é possível dizer que, para esses substantivos, o gênero está expresso tanto no nível sintático quanto na estrutura do vacábulo.

Em relação aos substantivos sobrecomuns (*testemunha, vítima, criança, pessoa*), a diferenciação de sexo não é possível por meio do

expediente sintático (ROCHA, 2008). O que está expresso no plano sintático é o gênero imanente do substantivo: *pessoa*, *vítima*, *testemunha* e *criança* são todos substantivos femininos, independentemente de se referirem a seres do sexo masculino ou feminino. Essa classe de substantivos obriga a seleção da forma feminina do determinante e o acréscimo da desinência de gênero feminina nos adjetivos. Nesses casos, o gênero pode ser deduzido do contexto em que os substantivos estão inseridos.

2.3.1 Elementos marcados pelo gênero no português

Nas seção anterior (seção 2. 3) mencionei que o gênero gramatical do português se realiza na língua por meio da concordância. Mencionei, também, que determinantes e adjetivos concordam obrigatoriamente em gênero com o nome substantivo que acompanham. Cabe aqui descrever de que forma essa concordância se manifesta nesses outros elementos que compõem o sintagma nominal.

Radford (1997) define como determinantes os itens que determinam propriedades referenciais dos nomes que acompanham. Assim sendo, são considerados determinantes os pronomes possessivos e pronomes demonstrativos, pronomes indefinidos (*meu/minha, aquele/aquela, este/esta, um/uma, algum/alguma, todo/toda*), numerais, artigos definidos (*o, a*).

No português, os determinantes devem concordar em gênero e número com os substantivos que acompanham. Muitos desses itens apresentam marcas de gênero em sua estrutura. Por exemplo, os pronomes possessivos *meu (s), nosso (s), teu (s), seu (s)* são selecionados pelos falantes do português para fins de concordância com um substantivo do gênero masculino, ao passo que *minha (s), nossa(s), tua(s), sua(s)* são selecionados para realizar a concordância com substantivos do gênero feminino.

Os pronomes indefinidos podem ou não ser variáveis quanto ao gênero. Em um sintagma nominal, o pronome *um* acompanha substantivos masculinos, e o pronome *uma* substantivos femininos. Conforme aponta Azeredo (2013, p.179), são variáveis em gênero os pronomes indefinidos: *um, algum, certo, determinado, muito, nenhum, outro, pouco, próprio, quanto, tanto, todo, vários*. São invariáveis quanto ao gênero os pronomes indefinidos *cada, demais, mais, menos, que*, e todos os pronomes que são núcleo de um sintagma nominal, *algo, alguém, nada, ninguém, outrem, quanto, que, quem, tudo*, bem como as locuções *o que* e *quem quer que*. Pronomes demonstrativos apresentam um comportamento semelhante aos pronomes indefinidos, isto é, podem ou não variar em gênero. São variáveis em gênero: *este (s), aquele (s)*. São invariáveis em gênero: *Isto, aquilo*.

No que se refere aos artigos definidos, o artigo masculino *o* é anteposto ao substantivo masculino; o artigo feminino *a* é anteposto ao substantivo feminino (CUNHA; CINTRA, 2001). Nesse sentido, os artigos definidos são variáveis em gênero e, por isso, são também marcadores de gênero no plano sintático.

Em relação aos numerais, apenas *um* e *dois* são variáveis em gênero: *um carro, uma casa; dois bancos, duas cadeiras*. Os demais são invariáveis. Por exemplo, *seis carros; seis casas, dez carros, dez casas*.

Os adjetivos do português são também marcados pelo gênero do substantivo que acompanham num processo de concordância. Os nomes adjetivos recebem a desinência de gênero *-a* para marcar o gênero feminino, e *-o* para marcar o gênero masculino. Enquanto os substantivos assumem obrigatoriamente um gênero (feminino ou masculino), os adjetivos são ora femininos (por exemplo: *clara, bonita, rica*), ora masculinos (por exemplo: *claro, bonito, rico*), ora neutros (por exemplo: *interessante, inteligente, grande*). Assim, o contraste de gênero feminino/ masculino dá-se para a totalidade dos nomes substantivos. Quanto aos nomes adjetivos, o contraste de gênero feminino/masculino dá-se apenas para um conjunto de adjetivos (*claro/clara*), os demais são neutros (*grande/interessante*), ou seja, não recebem informação de gênero nem mesmo no nível sintático (cf. VILLALVA, 2000, CÂMARA JR., 1970).

Fora do sintagma nominal, o termo que exerce a função de predicativo também recebe marcas de gênero na concordância (cf. NAME, 2002). Exemplos de concordância em gênero com o predicativo são:

- a) As crianças pareciam *envergonhadas* aos olhos do viajante;
- b) Todos acusaram-no de *desmotivado*.

Em suma, na língua portuguesa, determinantes e adjetivos concordam em gênero com os substantivos a que se referem. Esses determinantes e adjetivos, no entanto, são variáveis ou não em gênero. Quando variáveis, os adjetivos recebem a vogal átona final *-a* para concordar em gênero com substantivos femininos, ou a vogal átona final *-o* para concordar com substantivos masculinos. De forma semelhante, o predicativo do sujeito concorda em gênero com o substantivo a que se referem. Quando fazem referência a substantivos femininos, recebem a vogal átona final *-a*. Quando fazem referência a substantivos masculinos, recebem a vogal átona final *-o*.

Na próxima seção, resumo a discussão apresentada neste capítulo.

2.4 Resumo da discussão

Neste capítulo, vimos que todos os nomes substantivos do português apresentam obrigatoriamente um gênero. Esse pode ser um traço imanente, parte integrante do nome substantivo, ou pode ser variável. Em nomes que designam seres animados, o gênero pode ser imanente ou variável. Em um subconjunto de substantivos que designam entidades inanimadas, o traço só pode ser imanente. Os substantivos de referente animado, por sua vez, podem apresentar marcação de gênero em sua estrutura morfológica a qual expressa uma informação de ordem semântica: o sexo biológico do referente. Alguns autores consideram que esse subconjunto de substantivos flexionam em gênero. Em contrapartida, outros autores assumem que a formação do gênero feminino trata-se de um processo derivacional.

O Quadro 4, a seguir, adaptado de Name (2002, p. 32), resume a classificação dos nomes em português quanto à natureza do traço de gênero.

Quadro 3: Classificação dos nomes em português quanto à natureza do traço de gênero Segundo Name (2002).

Traço de gênero dos nomes		
animacidade	Opcionalidade	
	Intrínseco	Opcional
[-animado]	<i>mesa, livro</i>	-
[+animado]	<i>cônjuge, girafa</i>	<i>menino/menina</i>
	<i>dentista, colega</i>	

Apesar do gênero intrínseco aos substantivos do português que designam entidades inanimadas parecer um traço arbitrário (cf. NAME, 2002; TAFT; MEUNIER, 1998), neste capítulo, vimos que, para um grande subconjunto de substantivos, é possível estabelecer um relação entre os fonemas finais e seu gênero gramatical, permitindo que os

falantes do português façam certas generalizações. (cf. NAME, 2002; CORBETT, 1991; TAFT; MEUNIER, 1998; TUCKER; LAMBERT; RIGAULT; SEGALOWITZ, 1968). Em contrapartida, um grande conjunto de substantivos não apresenta em sua estrutura fonológica nenhuma pista de gênero, ou sua estrutura (terminação) contradiz a expectativa de seu gênero. Em ambos os casos, o gênero só está expresso na concordância por meio da marcação na estrutura dos determinantes e modificadores variáveis (adjetivos, predicativos). Com base nessas diferenças estruturais, os nomes substantivos foram agrupados em quatro categorias de formas quanto ao gênero: formas regulares (substantivos femininos terminados em *-a*, e masculinos terminados em *-o*), irregulares (substantivos femininos terminados em *-o*, e masculinos terminados em *-a*), transparentes (substantivos com terminações típicas do gênero feminino, como: *-agem, -ade, -ente, -ice*, e substantivos com terminações típicas do gênero masculino, como: *-or, -ume, -ema*) e opacas. As diferenças/semelhanças entre os processamentos dessas categorias de formas serão investigadas nos experimentos apresentados nos capítulos 4-6.

Vale lembrar e esclarecer que, embora este capítulo apresente a descrição do conceito de gênero do português tanto do referente animado [\pm humano] quanto do referente inanimado [- humano], nesta tese, detenho-me na análise do processamento do gênero dos substantivos de referente inanimado e no processamento da concordância com determinantes e adjetivos variáveis. A investigação do processamento dessas categorias de formas dos substantivos inanimados do português será guiada pelos pressupostos dos modelos de processamento morfológico apresentados no próximo capítulo (Capítulo 3).

CAPÍTULO 3

3 MODELOS DE PROCESSAMENTO MORFOLÓGICO

3.1 Introdução

A pesquisa no campo da Psicolinguística e Neurocognição da Linguagem com foco no processamento morfológico tem abordado, entre outros temas, a distinção entre o processamento e a representação mental de itens lexicais regulares e irregulares. Modelos de processamento morfológico discutem se itens lexicais regulares e irregulares são processados pelos mesmos ou por diferentes mecanismos (neuro)cognitivos. Esses modelos estão ancorados em dois arcabouços teóricos principais. São eles: os modelos teóricos de via dual (*dual-systems theories*) e os modelos teóricos de via única (*single-systems theories*).

Os modelos de via dual - ou dupla rota - (PINKER, 1991, 1998, 1999; ULLMAN, 2001a, 2001b, 2004, 2005) postulam que a linguagem humana compreende dois sistemas diferentes: o léxico mental e a gramática mental. Em contrapartida, os modelos de via única - ou rota única - (BATES; WULFECK, 1989; BATES; WULFECK, 1991; BATES; MacWHINNEY, 1989; MACWHINNEY, 2002, 2001, 2005; CHATER; MANNING, 2006; BYBEE, 1995; JOANISSE; SEIDENBERG, 1999) não admitem uma distinção neurocognitiva entre gramática mental e léxico mental. Para os pesquisadores dessa vertente, só há um sistema para o processamento da linguagem: um sistema de memória de natureza associativa que funciona como rede integrada de conexões neurais (BYBEE, 1995).

Os modelos de via dual representam o paradigma simbólico que entende a linguagem humana como um conhecimento representado por regras lógicas e manipulação de símbolos (cf. PINKER, 1991, 1998). Os modelos de via única, por sua vez, estão em consonância com o paradigma conexionista da linguagem que despensa a ideia de dualidade de sistemas cognitivos. Os conexionistas compreendem a linguagem humana como um entrelaçamento de diversos tipos de processamentos em vários níveis: desde o nível neural até o meio em que o falante está inserido.

Neste capítulo, apresento os principais construtos teóricos e hipóteses de representação e processamento dos modelos teóricos de via dual e via única mais relevantes para a investigação proposta nesta tese. Apresento, também, os resultados de estudos empíricos que testaram as hipóteses levantadas por esses modelos.

3.2 Pinker: *words and rules*

Na *introdução* deste capítulo mencionei que, segundo os modelos de via dual, a linguagem humana compreende dois mecanismos cognitivos distintos: o léxico mental e a gramática mental. No modelo proposto por Pinker (1991, 1998, 1999), o léxico mental seria um grande repositório de signos linguísticos (PINKER, 1998). A gramática mental, por sua vez, seria um sistema de natureza simbólica responsável pela computação de regras gramaticais combinatórias o qual permite que um número ilimitado de novas combinações seja processado.

Para Pinker, o léxico e a gramática são sistemas distintos, tanto do ponto de vista cognitivo quanto do ponto de vista neurológico. No entanto, apesar de os dois sistemas apresentarem um funcionamento interativo, eles também funcionam de modo competitivo. Por exemplo, se um determinado item lexical apresentar qualquer traço idiossincrático, então a gramática mental é inibida para que o item seja acessado e processado no léxico mental. Da mesma forma, se o item lexical exigir aplicação de regras de natureza computacional, o léxico mental é inibido.

Pinker propõe como forma de testar a sua proposta de arquitetura da linguagem contrastar o processamento de itens lexicais regulares e irregulares. Ele parte da ideia de que as regras que incidem sobre o radical de itens regulares são de natureza computacional e, por isso, itens lexicais regulares são processados *online* na gramática mental. Em contrapartida, itens irregulares, não passíveis de aplicação de regras, seriam armazenados e acessados em sua forma pronta no léxico mental.

Na visão dual de processamento, o fato de a gramática mental e do léxico mental serem sistemas distintos e exercerem funções distintas resulta em processamentos distintos. Portanto, há fatores que afetam o processamento do léxico mental, mas não afetam a gramática mental. De acordo com o modelo de via dual proposto por Pinker (1991, 1999), a frequência de uso não exerce qualquer influência no processamento e representação de itens regulares, ao passo que os irregulares, idiossincráticos em sua forma, são sensíveis à frequência de ocorrência na língua, já que são memorizados e acessados como uma unidade no léxico mental.

Pinker (1991, 1998) esclarece que há uma correlação massiva entre frequência de forma de superfície e irregularidade. Por outro lado,

há uma correlação igualmente massiva entre frequência de lexema e regularidade. Segundo o autor, 98% dos verbos infrequentes do inglês (frequência em torno de 1/2 itens por milhão) são regulares, ou seja, recebem o sufixo *-ed* de formação do passado. Em contrapartida, os itens mais frequentes são as formas de superfície (*went, spoke, was, were*) de verbos irregulares. A explicação para esse fenômeno, como veremos a seguir, é a mesma explicação de Bybee (1995). Se uma palavra é pouco usada pelos membros de uma comunidade, isto é, se uma palavra é infrequente, há uma tendência para que os falantes da língua decidam pela regularização de sua forma, visto que a forma derivada é de mais difícil acesso na memória em virtude de sua infrequência.

O léxico mental, visão dual de processamento, seria um sistema de memória responsável, também, por processos de natureza associativa entre itens da classe dos irregulares. Por exemplo, padrões estruturais em formas irregulares e semelhanças fonológicas (por exemplo, combinação de fonemas na estrutura silábica) também afetam o processamento da memória associativa. Assim como o sistema associativo proposto pelos modelos de via única, as entradas no léxico mental estão agrupadas com base em suas semelhanças semânticas, morfológicas, fonológicas e ortográficas. Portanto, as subregularidades existentes em itens irregulares (por exemplo, *blink/blank, drink/drunk, blow/blew, flow/flew*) estão a cargo do léxico mental.

Vários estudos sobre o tema mostraram efeitos de frequência e processos associativos incidindo apenas em um conjunto de itens irregulares (por exemplo, FLEISCHHAUER; CLAHSSEN, 2012; VAN DER LELY; ULLMAN, 2001a; PRADO; ULLMAN, 2009, PRASADA; PINKER, 1993) os quais representam evidências para os modelos híbridos de processamento. O estudo de Prasada e Pinker (1993) fornece evidências para essa hipótese ao mostrar que as generalizações feitas entre itens lexicais irregulares dependem do grau de similaridade entre eles, ao passo que a aplicação de regras independe dessa similaridade. A aplicação de regras de natureza computacional pode se estender a novos itens independentemente das semelhanças e independentemente da frequência de ocorrência do item. Assim, para os modelos de via dual, associações estruturais entre itens irregulares são funções desempenhadas pelo léxico mental e associações de regras estão a cargo da gramática mental.

Apresento, a seguir, um outro modelo de via dual, de natureza neurocognitiva, cujos pressupostos teóricos estão fundamentos na proposta de Prasada e Pinker (1993) e Pinker (1991, 1998, 1999).

3.3 Ullman: *Modelo Declarativo/Procedural*

O Modelo Declarativo/Procedural (doravante MDP) proposto por Ullman (ULLMAN, 2001a; 2001b; 2004; 2005), assim como o modelo proposto por Pinker (1991, 1998, 1999), parte da premissa de que a linguagem humana compreende dois sistemas cognitivos distintos: o léxico mental e a gramática mental. Contudo, o MDP amplia as discussões em relação a alguns aspectos: a organização neurofuncional do léxico e da gramática mental bem como as funções desempenhadas por ambos os sistemas.

O MDP postula que os substratos neurais do léxico mental e da gramática mental não são específicos da linguagem, mas lidam com outras funções, portanto, são de domínio-geral (*domain-general*). De acordo com esse modelo, os substratos neurais do léxico mental e da gramática mental são sistemas separados, mas interdependentes que funcionam cooperativamente e competitivamente. Ullman (2001b) afirma que as estruturas do lobo temporal medial, como o hipocampo e estruturas relacionadas, as quais estão conectadas às regiões temporais e parietais do córtex cerebral, subjazem à memória declarativa. Os componentes do lobo medial temporal são especializados na consolidação de novas memórias e memória relativa a fatos e eventos. Segundo Ullman (2001b), as estruturas do lobo medial temporal são responsáveis pelo processamento de informações aprendidas conscientemente. O léxico mental, segundo o MDP, depende das estruturas do sistema de memória declarativa (cf. ULLMAN, 2001a, 2001b, 2005).

As estruturas frontais e os gânglios da base subjazem à memória procedural. Segundo o MDP (cf. ULLMAN, 2001b), os gânglios da base estão conectados às estruturas do córtex frontal via o tálamo. Esse circuito é especializado na aprendizagem de habilidades que requerem a aplicação de regras de ordem sequencial como a aplicação de regras morfosintáticas. Assim, a memória procedural subjaz à gramática mental. Esse sistema está relacionado, também, ao processamento de habilidades de controle motor e ao processamento de hábitos rotineiros como, por exemplo, dirigir um carro, andar de bicicleta, tocar violão. De acordo com o MDP, o conhecimento do sistema procedural é de natureza implícita e o conhecimento retido na memória declarativa é de natureza explícita/associativa (ULLMAN *et al.*, 1997).

O léxico mental seria composto por pares arbitrários de som/significado. Esses pares arbitrários podem ser simples palavras

(como *gato*, *cat*), formas verbais irregulares (*era*, *foi*, *went*) ou estruturas mais complexas, como palavras compostas (*mesa-redonda* ou *guarda-chuva*). Por outro lado, a gramática mental compreende o conhecimento implícito de regras que governam a linguagem, entre elas, as regras morfológicas de flexão regular (como a flexão verbal, flexão de gênero ou flexão de número) além da sintaxe, isto é, a combinação hierárquica e sequencial das entradas lexicais de modo a fazerem sentido em uma sentença. Na versão do MDP formulada em Ullman, Estabrooke, Steinhauer, Brovetto, Pancheva, Ozawa, Mordecai and Maki (2002) and Ullman (2004), itens lexicais irregulares estão armazenados na memória declarativa, ao passo que itens regulares são processados na memória procedural responsável pela codificação de regras e pela decomposição dos itens regulares em constituintes em processos de compreensão e pela concatenação de uma base a um afixo em processos de produção.

Em relação à L2, o modelo argumenta que adultos, aprendizes tardios com baixa proficiência em uma segunda língua, podem depender mais da memória declarativa do que da memória procedural uma vez que o conhecimento sintático complexo - que deveria ser processado pela memória procedural - é armazenado na memória declarativa como pedaços congelados da língua (*chunks*). Assim, falantes adultos com baixa proficiência podem armazenar todo o conhecimento da L2 na memória declarativa, por não serem capazes de ativar o sistema procedural nos estágios iniciais de aprendizagem. Por outro lado, falantes experientes na L2 podem usar os dois sistemas de memória assim como falantes da L1 em função do tempo de exposição e prática na segunda língua.

Apesar de considerar a sintaxe como um componente importante da memória procedural, o MDP, assim como o modelo de via dual proposto por Pinker (1998), testa suas hipóteses a partir da morfologia flexional - principalmente, a partir da morfologia flexional verbal do passado do inglês. A morfologia verbal tem sido um tópico de pesquisa recorrente de estudos que visam testar as hipóteses dos modelos de via dual, já que a morfologia verbal do pretérito passado do inglês permite investigar a existência de ambos os sistemas de memória.

De acordo com a visão dual de processamento, os verbos regulares, ou seja, verbos que seguem um paradigma flexional totalmente previsível (como a adição do morfema flexional *-ed* para formação do passado em inglês: *walk-walked*; *remember-remembered*) seriam processados pela memória procedural, pois aplica-se a esses o componente computacional. Em contrapartida, os verbos irregulares, imprevisíveis em sua forma, seriam processados como uma nova entrada

lexical na memória declarativa, pois, a eles, o componente computacional não pode ser aplicado (por exemplo, *think-thought, seek-sought*).

O MDP prevê, também, diferenças individuais entre o processamento do conhecimento de tipo lexical/semântico e o conhecimento de tipo gramatical. Segundo o modelo, enquanto as mulheres seriam mais hábeis em tarefas relativas à memória declarativa, por exemplo, em tarefas relativas à memorização de palavras, os homens seriam melhores em tarefas específicas da memória procedural. Segundo Ullman *et al.* (2002) Ullman (2005), as diferenças entre os sexos são decorrentes dos níveis do hormônio estrogênio no cérebro feminino e masculino.

Um método que vem sendo muito utilizado para testar as hipóteses do MDP é a análise dos efeitos de frequência (do inglês *frequency effects*). Esse método é baseado na hipótese de que formas lexicais frequentemente encontradas na língua são mais rapidamente acessadas na memória e, por isso, o tempo de reação aos itens frequentes é menor quando comparado ao tempo de reação a itens infrequentes. Vários estudos focalizando no contraste existente entre formas regulares e irregulares analisaram os efeitos de frequência em itens regulares e irregulares.

Os resultados de vários estudos (por exemplo, ULLMAN, 1999; BOWDEN, 2007; MORGAN-SHORT *et al.*, 2010) revelam que formas irregulares frequentes são mais rapidamente acessadas na memória que formas irregulares infrequentes. Portanto, os resultados mostram que, sobre itens lexicais irregulares, incidem efeitos de frequência. Por outro lado, os resultados desses estudos mostraram que itens regulares frequentes e infrequentes não suscitam efeitos de frequência e, por isso, são supostamente gerados por meio de um processo computacional de concatenação de uma base a um afixo (*walk + ed = walked*).

Em resumo, o MDP é um modelo de via dual e postula que os mecanismos responsáveis pelo léxico e pela gramática mental são de domínio-geral em oposição aos modelo de via dual clássicos (por exemplo, PINKER, 1998; CHOMSKY, 1981, 1995) que postulam sistemas neurocognitivos específicos para o processamento do conhecimento linguístico. O MDP apresenta, ainda, hipóteses acerca da aquisição e processamento da L2. Esse modelo preconiza que, na aquisição da L2 por adultos, o processamento pode ocorrer principalmente no sistema de memória declarativa nos estágios iniciais de aprendizagem. Esse modelo prevê, também, diferenças cognitivas entre homens e mulheres nas habilidades referentes a um e outro

sistema. O MDP assume uma distinção (neuro)cognitiva entre o conhecimento de tipo lexical/semântico e conhecimento de tipo gramatical: enquanto os itens lexicais irregulares/ idiossincráticos em sua forma, são armazenados em sua forma pronta e processados no sistema de memória declarativa, os itens regulares/composicionais são processados no sistema de memória procedural.

A seguir, descrevo resultados de estudos empíricos que trazem evidências para a visão dual de processamento.

3.3.1 Evidências: Modelos de via dual

Em um estudo seminal, Oldfield e Wingfield (1965) observaram que figuras de palavras frequentes eram reconhecidas mais rapidamente que figuras de palavras menos frequentes. Esse estudo foi replicado por outros pesquisadores (JESCHENIAK *et al.* 1994; LEVELT; PRAAMSTRA; MEYER; HELENIUS; SALMELIN, 1998), utilizando diferentes tarefas linguísticas. Os resultados obtidos evidenciaram que palavras mais frequentes são mais rapidamente acessadas na memória que palavras infrequentes. A partir de então, a correlação existente entre a frequência de uso das palavras e velocidade de acesso lexical na memória é referida, por grande parte dos estudos, como os efeitos de frequência (cf. BIEN, 2007).

Os modelos de via dual partem dessa correlação para testar a hipótese da dissociação neurocognitiva entre o léxico e a gramática mental. Como já descrito, de acordo com a perspectiva dual, formas irregulares são armazenadas e processadas no léxico mental, um sistema de memória de natureza associativa sensível à frequência dos itens lexicais. Em contrapartida, itens regulares, frequentes ou infrequentes, são processados na gramática mental, um sistema responsável pelo processamento de conhecimento implícito de natureza computacional. Portanto, segundo a perspectiva dual de processamento, efeitos de frequência não afetam o processamento dos itens regulares, mas afetam o processamento de itens irregulares (cf. ULLMAN, 1999; PINKER; ULLMAN, 2002).

Os primeiros estudos nesse sentido foram realizados por Forster e Chambers (1973), Rubenstein, Garfield e Milliken (1970) e Seidenberg e Bruck (1990) e se concentraram na morfologia verbal do passado do inglês. Os resultados demonstraram que falantes do inglês como L1 produziam o passado dos verbos de forma mais rápida para itens

irregulares muito frequentes na língua do que para itens irregulares infrequentes. No entanto, efeitos de frequência não foram encontrados entre os itens regulares (itens que exigiam o acréscimo do morfema flexional *-ed* ao radical do verbo) cuja flexão era realizada aplicando-se a regra de flexão padrão. Como consequência da série de investigações com foco na morfologia verbal, posteriormente, vários outros estudos surgiram com o objetivo de replicar os resultados para outras línguas (por exemplo, CLAHSSEN, ROTHWEILER; WOEST, 1992; MARCUS; BRINKMANN; CLAHSSEN; WIESE; PINKER, 1995).

Além dos efeitos de frequência, outros métodos também se mostraram adequados para investigar a dissociação cognitiva entre léxico mental e gramática mental. Por exemplo, Clahsen, Eisenbeiss e Sonnenstuhl (1997) estudaram a flexão verbal do particípio de verbos do alemão e a flexão de número da mesma língua. Os pesquisadores usaram o método conhecido como *priming* tanto para os verbos quanto para os nomes. Esse método consiste na avaliação do efeito que um estímulo antecedente (*prime*) exerce sobre um estímulo apresentado posteriormente (alvo). Pesquisadores que fazem uso desse método pressupõem que se o *prime* ativa o acesso da palavra alvo. Se o *prime* e o alvo estiverem conectados no léxico mental por meio de certos traços estruturais compartilhados, então o alvo é acessado mais rapidamente, conseqüentemente, o tempo de reação numa tarefa de decisão lexical será menor.

No estudo de Clahsen, Eisenbeiss e Sonnenstuhl (1997), os participantes escutaram itens regulares do alemão (particípio em *-t* e plural em *-s*) e irregulares (particípio em *-n* e plural em *-er*). Em seguida, os alvos das palavras escutadas apareceram visualmente em uma tela de computador. Os participantes foram solicitados a produzirem o passado e o plural de pseudopalavras e palavras reais logo após a apresentação do *prime*, ou seja, do item lexical que fornece informações para a produção do alvo. Os resultados desse estudo demonstraram um efeito de *priming* total para itens regulares, mas parcial para itens irregulares. Além disso, os participantes foram mais velozes para itens regulares que para itens irregulares. Com base nesses resultados, os autores concluíram que itens irregulares são armazenados individualmente no léxico mental, ao passo que itens regulares estão em conexão com outras formas regulares.

Prasada e Pinker (1993) encontram mais evidências a favor dos pressupostos dos modelos de via dual. A investigação teve como objetivo testar as hipóteses dos modelos de via única e via dual para o

processamento de itens regulares e padrões fonológicos de itens irregulares. Para tanto, os pesquisadores implementaram três experimentos e três simulações usando redes neurais. Os resultados revelaram que as generalizações feitas com verbos irregulares dependem do nível de similaridade dos padrões estruturais (fonológicos) dos itens, ao passo que a aplicação da regra do passado do inglês independe da similaridade dos vocábulos. Ullman (1999) obteve resultados semelhantes. O estudo empregou um teste de julgamento gramatical de sentenças com verbos regulares e irregulares contendo padrões fonológicos em sua estrutura como *sing*, *sang*, *drink*, *drank*. Os resultados mostraram uma correlação entre frequência de ocorrência e o número de acertos para itens irregulares, ao passo que, entre os itens regulares, os resultados não revelaram nenhuma correlação entre frequência de uso e acertos. Os resultados foram interpretados pelo pesquisador como mais uma evidência da dualidade de processamento.

Outros estudos que merecem destaque são os de Penke, Janssen e Krause (1999) e Van Der Lely e Ullman (2001) os quais investigaram pacientes afásicos. Penke, Janssen e Krause (1999) estudaram o processamento de itens regulares e irregulares em 11 afásicos, lesionados na área de Broca e sofrendo de agramatismo, com o objetivo de verificar se o processamento de itens regulares e irregulares estaria comprometido nesses sujeitos. Os resultados indicaram que o processamento de itens irregulares foi afetado, ao passo que o de itens regulares não, sugerindo a dissociação cognitiva dos sistemas de processamento. Em acréscimo, de forma análoga aos estudos que investigaram adultos saudáveis, efeitos de frequência associados aos erros cometidos pelos pacientes foram encontrados nos itens irregulares. Em contrapartida, efeitos de frequência não foram encontrados para itens regulares.

Van Der Lely e Ullman (2001) compararam o desempenho de crianças com distúrbios no desenvolvimento da linguagem e crianças saudáveis na produção do passado de verbos regulares e irregulares do inglês. As crianças saudáveis mostraram desempenho superior a crianças com distúrbios. Os resultados revelaram, também, uma velocidade de resposta superior para a produção dos verbos regulares em comparação a produção dos verbos irregulares. Novamente, esses resultados sugerem a dissociação dos processamentos de itens regulares e irregulares.

Recentemente, os resultados do estudo de Fleischhauer e Clahsen (2012) trouxeram mais evidências para a hipótese da dualidade de processamento. Nesse estudo, os pesquisadores compararam o desempenho de adultos e crianças em uma tarefa de produção do

particípio passado dos verbos do alemão. Os resultados obtidos mostram uma vantagem na velocidade de resposta, em ambos os grupos de participantes, para os itens frequentes sobre as formas de baixa frequência para particípios irregulares flexionados. Para itens regulares, no entanto, uma desvantagem significativa de frequência alta sobre formas de baixa frequência foi encontrada para as crianças, mas não para adultos. Apesar dos surpreendentes resultados encontrados para o grupo de crianças, os autores interpretam os resultados como mais uma evidência para a dualidade de sistemas cognitivos.

Apesar das evidências que dão suporte aos modelos de via dual, outros estudos apresentam indícios a favor dos modelos de via única. Apresento, na próxima seção, as premissas dos modelos de via única, em especial, do *Network Model* (BYBEE, 1985, 1995). Na sequência, descrevo os resultados de alguns estudos que fornecem evidências para as hipóteses levantadas por esses modelos.

3.4 Bybee: *Network Model*

A visão dual de processamento e representação da morfologia tem sido veementemente contestada pelos modelos conexionistas e por outros modelos de via única como, por exemplo, o *Network Model* proposto por Bybee (1995). Em oposição aos modelos de via dual, esse modelo rejeita a ideia de dissociação cognitiva dos sistemas de memória para o processamento e representação de itens lexicais regulares e irregulares. Para o *Network Model*, itens regulares e irregulares estão representados em uma memória de caráter associativo e por ela são processados.

Segundo Bybee (1995), os itens lexicais distinguem-se quanto à força de sua representação no léxico mental. A força de representação de um item lexical é determinada pela relação entre suas propriedades morfológicas/fonológicas e sua frequência de uso. Se a forma de superfície (*token frequency*) de um item lexical é altamente frequente, então essa forma é mais resistente à mudança e, por isso, é mais propensa a adquirir e manter sua independência no sistema de memória dos falantes. No entanto, se uma forma irregular apresenta baixa frequência, então essa forma estará mais propensa à regularização, já que não é de fácil acesso na memória. De acordo com a Bybee (1995), essa relação explica porque itens lexicais irregulares apresentam uma frequência de forma de superfície maior do que a frequência de forma de superfície dos itens regulares.

O *Network Model* prevê que a frequência incide tanto sobre itens regulares quanto sobre itens irregulares. Assim, se um item irregular é frequente, então a memória associativa, que funciona com uma rede de conexões (itens que compartilham traços estão em conexão), o toma como modelo para associar suas propriedades mórficas/fonológicas com outras formas que compartilham traços semelhantes (por exemplo, *fly-flew- blow-blew*). Isso significa dizer que itens lexicais que compartilham traços estão interligados no léxico mental. Para o *Network Model*, as conexões lexicais também podem variar em sua força conforme o tipo e número de traços compartilhados. Dessa forma, a quantidade de traços compartilhados também representa um fator que contribui para força de representação do item na memória associativa. Quanto mais traços compartilhados, maior a conexão entre os itens, consequentemente, maior será a força das associações as quais, por sua vez, podem ser estendidas a novos itens. Em acréscimo, as conexões lexicais podem ser afetadas pela força lexical de um item. Palavras que apresentam uma frequência de forma alta têm maior autonomia lexical, consequentemente, essas palavras formam conexões mais fracas com outras palavras. Bybee explica que a forma *went* (passado do verbo *to go* do inglês), por exemplo, é altamente frequente e, por isso, adquiriu autonomia lexical suficiente para não estar associada nem a sua forma de base (*to go*), nem a outras palavras que compartilham traços fonológicos semelhantes, como o verbo *to wend*. Com base nessa hipótese, Bybee sugere que itens altamente frequentes, que já adquiriram uma autonomia lexical, são aprendidos autonomamente. Em contrapartida, itens menos frequentes são melhor aprendidos por associações com outros itens já existentes.

Bybee (1995) apresenta evidências para essa argumentação. A autora descreve um estudo que aplicou a técnica de *-priming*⁷ usando verbos irregulares frequentes e infrequentes antes de apresentar uma lista de pseudoverbos que lembravam os verbos naturais apresentados. Nesse experimento, os participantes produziram mais pseudoverbos que se assemelhavam aos itens irregulares infrequentes do que os que se assemelhavam aos itens irregulares frequentes.

Outra questão abordada pelo modelo diz respeito aos tipos de conexões que se pode formar. Se um conjunto de palavras apresenta padrões semelhantes de conexões semânticas e fonológicas, elas reforçam umas às outras, resultando em generalizações que podem ser

⁷ Confira explicação desse metodologia na seção 3.2.1

estendidas a novos itens. Essas generalizações são descritas pelo modelo como esquemas (*schemas*). A probabilidade de um esquema ser estendido a um novo item depende de suas propriedades e de sua força. Se um esquema é forte, então a produtividade é alta; se o esquema é baixo, sua produtividade é baixa. A produtividade do esquema, por sua vez, é baseada na frequência dos padrões que ele representa. Quanto maior a frequência dos padrões descritos no esquema, maior é a chance do esquema ser estendido a novos itens.

Nas palavras da autora:

Formas de baixa frequência de *token* serão aprendidas mais facilmente se elas puderem estar relacionadas com outras formas armazenadas. Se o *input* contém uma grande quantidade de itens distintos que compartilham um afixo, estas palavras estarão conectadas umas às outras no léxico, e a existência do afixo emergirá [...]. Quanto mais formas que contém o afixo, mais forte é a representação do afixo. Quanto mais forte a representação do afixo, mais fácil ele será acessado quando uma nova palavra precisa ser flexionada e maior é a probabilidade do afixo ser produtivo. (BYBEE, 1995, p.434, tradução nossa).⁸

O *Network Model* propõe dois tipos de esquemas: o esquema *source-oriented* e o esquema *product-oriented*. O esquema *source-oriented* compreende generalizações que envolvem formas básicas como a raiz dos vocábulos. Os modelos de via-dual entendem essas generalizações como aplicação de regras de natureza computacional com em *walk-walked*. Já as generalizações *product-oriented* agem sobre um conjunto de formas complexas ou derivadas que compartilham traços semelhantes como em *fly-flew-blow-blew*. Assim, mesmo que uma palavra não possa ser decomposta em seus constituintes, se a

⁸ Forms with lower token frequency will be learned more easily if they can be related to other stored forms. If the input contains a large number of distinct items which share an affix, these words will be related to one another in the lexicon, and the existence of the affix will emerge [...]. The more forms that bear the affix, the stronger the representation of that affix. The stronger the representation of the affix, the easier it will be to access when a new word needs to be inflected, and the greater likelihood that affix will be productive.

produtividade do esquema a que ela pertence é alta, generalizações com base em seus padrões morfológicos/fonológicos podem ser estendidas a novos itens.

Sendo assim, de acordo com o *Network Model*, tanto itens regulares quanto itens irregulares sofrem efeitos de frequência, pois são processados por uma memória associativa. A frequência é entendida com um aspecto que influencia os dois tipos de esquema de associações: quanto maior a frequência descrita no esquema, mais associações serão feitas com outros itens.

3.4.1 Evidências: modelos de via única

Os modelos de via única assumem que formas regulares e irregulares são processadas pelos mesmos mecanismos neurocognitivos e, por isso, ambas as categorias de forma estão sujeitas a efeitos de frequência e processos de natureza associativa.

O estudo de Alegre e Gordon (1999) representa um exemplo clássico a favor dessa hipótese. Os resultados do estudo revelaram que formas regulares altamente frequentes e itens irregulares podem ser armazenados em sua forma pronta. Em cinco experimentos, ao manipular itens lexicais com diferentes graus de frequência, os pesquisadores mostraram que o armazenamento de formas regulares é possível se a frequência do item regular for superior a um determinado limite, isto é, se o item regular é altamente frequente, então é possível que este item esteja armazenado. Por outro lado, efeitos de frequência não foram encontrados nos itens com frequência inferior a esse limite, indicando a decomposição morfológica do vocábulo. De modo geral, os resultados desse estudo apressentam evidências a favor dos modelos de via única que prevêm efeitos de frequência para itens regulares.

Resultados semelhantes já tinham sido obtidos anteriormente por Baayens e colaboradores (cf. BAAYEN *et al.*, 1997) ao estudarem a morfologia verbal e de número do holandês. Nesse estudo, a velocidade de resposta foi mensurada em uma tarefa de decisão lexical. Os resultados revelaram que há armazenamento para itens regulares altamente frequentes tanto para verbos quanto para nomes. Em um estudo subsequente, Baayen, Tweedie e Schreuder (2002) investigaram novamente a morfologia verbal e flexional de número no holandês. Da mesma forma, efeitos de frequência para itens lexicais regulares foram encontrados. Os autores atribuíram o armazenamento de itens regulares a fatores relacionados à frequência do item e a sua complexidade

morfológica.

Baayen (2003) replicou o estudo de 1997 usando uma modalidade auditiva da tarefa visual empregada na pesquisa anterior. Os resultados encontrados para a modalidade auditiva foram semelhantes aos resultados obtidos no experimento visual: itens regulares muito frequentes foram acessados com maior velocidade na memória que itens regulares infrequentes. Outros estudos que mostram efeitos de frequência para regulares são: Taft (1979) e Baayen, Schreuder, De Jong and Krott, (2002). Recentemente, Tabak, Schreuder e Baayen (2010), em quatro experimentos comportamentais usando uma tarefa de nomeação de figuras, também encontraram efeitos de frequência na produção de verbos regulares do holandês.

Além dos estudos comportamentais, os modelos conexionistas também fornecem evidências para os pressupostos dos modelos de via única. Redes neurais artificiais foram capazes de produzir o passado de itens regulares e itens irregulares com um alto grau de precisão com base nos padrões fonológicos presentes nos itens. O modelo de Plunkett e Marchman (1993), por exemplo, simula o processamento do passado de verbos regulares e irregulares do inglês, usando uma rede neural artificial. Em vários pontos do treinamento, a rede neural foi incrementada por meio da adição de vetores (verbos regulares e irregulares) ao arquivo de treino. As simulações mostraram que a rede é sensível às regularidades bem como às subregularidades morfofonológicas de verbos irregulares. Os erros nas simulações da rede foram equiparados a erros produzidos por crianças na fase de aprendizagem da língua materna.

O modelo conexionista implementado por Rumelhart e McClelland (1989) simulou as três fases do processo de aquisição do inglês por crianças. Na simulação da fase 1, o modelo reproduziu os estágios iniciais de aquisição do passado do inglês. Nesse estágio, um número reduzido de verbos regulares e irregulares coexistem no léxico mental das crianças. Esses verbos são todos memorizados, já que padrões estruturais dos verbos ainda não foram detectados. Assim como observado em crianças, o modelo conexionista não revelou diferenças na simulação da flexão de verbos regulares e irregulares. A simulação da fase 2 revelou que o modelo regularizava verbos irregulares, reproduzindo novamente um padrão observado nas fases seguintes do processo de aquisição da morfologia flexional verbal do passado do inglês por crianças. A simulação da fase 3 mostrou que o modelo é capaz de reconhecer os padrões estruturais de verbos regulares e irregulares nunca apresentados anteriormente à rede. Esses resultados

sugerem que o modelo foi capaz de reproduzir o comportamento de crianças durante as fases de aquisição da morfologia verbal do inglês. Sugerem, ainda, a sensibilidade do modelos conexionistas em relação aos padrões de regularidade bem como subregularidades de verbos do passado do inglês. Esses resultados evidenciam que o processamento tanto itens regulares quanto itens irregulares são passíveis de serem processados por um único mecanismo de caráter associativo, conforme preconizam os modelos de via única.

3.5 Resumo da discussão

Estudos empíricos que testaram as hipóteses de processamento morfológico levantadas pelos modelos de via dual e pelos modelos de via única concentraram-se, principalmente, na morfologia flexional verbal e, em menor proporção, na morfologia flexional de número. Como vimos, os resultados desses estudos são controversos: alguns resultados dão suporte aos pressupostos dos modelos de via dual, outros, aos modelos de via única. Tendo em vista as hipóteses de processamento e representação de itens regulares e irregulares pelos modelos de processamento, a questões que se tornam relevantes para esta tese são:

- 1) O processamento do gênero gramatical dos substantivos inanimados do português seria melhor representado por um sistema dual ou por um único sistema de memória associativa?
- 2) Em um processo de compreensão, os substantivos inanimados do português regulares em sua forma seriam decompostos em constituintes menores pelos falantes da língua para o processamento do seu gênero?
- 3) Substantivos transparentes, opacos e irregulares em forma seriam extraídos como uma unidade do léxico mental para o processamento de seu gênero?
- 4) Para o processamento do gênero gramatical, formas regulares entrariam em um processo associativo?

De fato, é possível encontrar na literatura estudos que investigaram a dualidade *versus* a singularidade do processamento do gênero gramatical de substantivos inanimados em várias línguas do mundo. Assim como os estudos com foco na morfologia verbal e de número discutem se formas regulares e irregulares recrutam os mesmos ou diferentes mecanismos neurocognitivos, estudos sobre o processamento do gênero gramatical de substantivos inanimados

discutem se o gênero de formas regulares e irregulares quanto ao gênero é processado por uma única ou por uma dupla rota. Alguns dos estudos mais relevantes sobre o tema serão descritos no próximo capítulo (Capítulo 4), na Parte II desta tese.

PARTE II: EXPERIMENTOS

PARTE II

PARTE II

A Parte II⁹ desta tese inicia com o Capítulo 4 em que apresento a metodologia e resultados de três experimentos comportamentais que tiveram como objetivo geral entender se o processamento do gênero gramatical de substantivos inanimados do português é governado por uma dupla ou por uma única rota. Como ponto de partida, decidi conduzir estudos comportamentais para que os resultados obtidos pudessem ser comparados aos resultados apresentados na literatura com foco no processamento do gênero gramatical em várias línguas do mundo. Em seguida, no Capítulo 5, apresento a metodologia empregada na construção de um modelo conexionista que foi utilizado na simulação do processamento do gênero de substantivos inanimados do português. O propósito dessas simulações foi investigar se um modelo conexionista de via única é capaz de reproduzir certos padrões de processamento observados nos resultados dos experimentos conduzidos com falantes do português apresentados no Capítulo 4. Na sequência, no Capítulo 6, apresento um experimento eletrofisiológico que teve como objetivo investigar se o processamento da concordância de gênero com substantivos inanimados regulares suscita os mesmos sinais eletrofisiológicos do processamento da concordância com substantivos das demais formas (opacas, transparentes e irregulares). Em todos os capítulos, reviso, primeiramente, estudos empíricos com foco no processamento do gênero gramatical em outras línguas que utilizaram métodos e procedimentos semelhantes. A minha proposta não é oferecer uma revisão exaustiva da literatura, mas, antes, apresentar estudos que são referência para o debate em torno da dualidade vs. singularidade do processamento da categoria gramatical de gênero nas línguas do mundo. Os projeto de pesquisa que inclui os experimentos apresentados no Capítulo 4 desta tese foi aprovado pelo Comitê de Ética em Pesquisa com Seres Humanos da Universidade Federal de Santa Catarina: Parecer número 359.055, aprovação número 16623613.8.0000.0121. O experimento eletrofisiológico apresentado no Capítulo 6 foi conduzido nas instalações do *Max Planck Institute for Psycholinguistics* em

⁹ Os *scripts* em *Linguagem R* criados para o pré-processamento e análises estatísticas dos dados apresentados no Capítulo 4 e Capítulo 6 bem como os *scripts* em *Linguagem R* das simulações computacionais conexionistas apresentadas no Capítulo 5 podem ser solicitados através do email: natcarol2000@gmail.com

PARTE II

Nijmegen, Países Baixos e, por isso, está sob regulamentação do Comitê de Ética em pesquisa da *Radboud Universiteit*.

CAPÍTULO 4

4 PROCESSAMENTO DO GÊNERO GRAMATICAL DO PORTUGUÊS: COMPUTAÇÃO DE REGRAS OU ASSOCIAÇÃO DE PADRÕES?

4.1 Introdução

O gênero gramatical, em especial, o gênero dos nomes substantivos que designam entidades inanimadas (tais como: *casa*, *razão*, *sol*), tem sido tópico recorrente de investigação entre os pesquisadores do campo da Psicolinguística. Em virtude do gênero dessa classe de substantivos não apresentar uma correlação com a noção de sexo, discute-se se o processamento do gênero de substantivos regulares ativa um mecanismo computacional ou se, todas os substantivos, independentemente de sua forma, são extraídos da memória e o processamento do gênero se realiza com base em associações entre traços fonológicos compartilhados pelas formas.

A ideia da ativação de um mecanismo computacional para o processamento de formas regulares quanto ao gênero está pautada no fato de haver, em muitas línguas do mundo, uma relação entre estrutura, principalmente terminação, e categorias de gênero (cf. CORBETT, 1991). Para citar alguns exemplos, tomemos como referência as línguas românicas como o espanhol, português e francês. No português e espanhol, é possível estabelecer uma relação entre vogal átona final *a* e o gênero feminino, e vogal átona final *o* e o gênero masculino. Nessas línguas, grande parte dos vocábulos terminados em *-a* são femininos, e terminados em *-o* são masculinos. No francês, a terminação *ie* (*folie*), por exemplo, está em estreita relação com o gênero feminino, e a terminação *-t* (*tricot*) com o gênero masculino (cf. TAFT; MEUNIER, 1998). A relação entre estrutura e gênero está presente, também, em línguas não românicas como o alemão e hebraico (cf. GOLLAN; FROST, 2001; HOHFELD, 2006).

Tendo em vista relação entre forma e gênero, alguns pesquisadores levantam a hipótese de que o gênero gramatical é processado por um mecanismo de natureza computacional responsável pela computação da regra da terminação. Segundo Corbett (1991), a regra da terminação seria uma regra fonológica de atribuição de gênero: se o substantivo termina no fonema *x*, então seu gênero é *y*. Corbett (1991, p.7) advoga a favor dessa posição ancorado em três argumentos: i) lembrar o gênero de palavra por palavra envolveria uma sobrecarga na memória dos falantes o que, ao contrário do que se percebe, acarretaria

em várias erros de concordância; ii) empréstimos de outras línguas adquirem, via de regra, um gênero, sugerindo um mecanismo para atribuição de gênero e não de memorização do gênero; iii) quando submetidos a pseudopalavras, falantes nativos de uma língua atribuem gênero com um alto grau de consistência.

De fato, é possível encontrar na literatura sobre o processamento do gênero gramatical evidências que dão suporte à hipótese de um mecanismo responsável pela computação da regra da terminação. Em um estudo seminal, Tucker, Lambert, Rigault e Segalowitz (1968) apresentaram a falantes nativos do francês pseudopalavras construídas com várias terminações típicas do gênero feminino e masculino da língua francesa. Os pesquisadores pediram que 400 sujeitos atribuíssem gênero aos itens. Os resultados mostraram que o gênero assinalado às pseudopalavras tinha uma correlação direta com a frequência de ocorrência da terminação e sua associação com uma das categorias: feminino ou masculino. Essa mesma associação foi encontrada em estudos posteriores (TUCKER; LAMBERT; RIGAULT, 1977). Com base nesses resultados, os pesquisadores concluíram que a atribuição de gênero em francês é governada por um conjunto de regras as quais foram internalizadas pelos falantes ao longo do processo de aquisição da língua. Resultados semelhantes foram obtidos para o alemão. Para essa língua, Hohlfeld (2006) mostrou que os falantes nativos computam o gênero (feminino, masculino ou neutro) com base nas marcas de gênero presentes na estrutura fonológica dos vocábulos.

Embora os resultados dos estudos aludidos indiquem a existência de um mecanismo responsável pela computação da regra da terminação, cabe refletir o que acontece quando o item não apresenta uma marca de gênero em sua estrutura ortográfica ou fonológica. Posto de outra forma: O que acontece quando o acesso ao sistema computacional é bloqueado? O estudo implementado por Gollan e Frost (2001) para o hebraico sugere um processamento em duas rotas. Nesse estudo, os pesquisadores conduziram dois experimentos. No primeiro experimento, os participantes foram requisitados a atribuir gênero a substantivos do hebraico contendo ou não sufixos indicativos de gênero (feminino ou masculino). No segundo experimento, apenas substantivos não marcados quanto ao gênero foram apresentados aos falantes: femininos irregulares e masculinos (em hebraico, apenas nomes femininos apresentam sufixo de gênero). Vale ressaltar que, no segundo experimento, apesar de apenas nomes não marcados terem sido apresentados aos falantes, o fator regularidade ainda estava presente, já que, no hebraico, nomes não marcados quanto ao gênero são masculinos. Em ambos os experimentos,

os resultados revelaram efeitos da regularidade: falantes foram mais rápidos para formas regulares (formas masculinas ou femininas contendo sufixos indicativos do gênero) do que para não marcadas (formas femininas irregulares).

Com base nesses achados, os pesquisadores sugerem dois mecanismos distintos para o processamento do gênero: um mecanismo responsável pelo processamento do gênero de regularidades fonológicas, isto é, itens cuja terminação fornece pistas para a atribuição de gênero e, além desse, um mecanismo responsável por formas não marcadas pelo gênero. Segundo os autores, nessa rota, o gênero não seria acessado no nível da forma, mas como uma informação abstrata incorporada ao lema dos itens lexicais.

Os resultados de estudos recentes, como o de Olívia *et al.*(2013) para o espanhol, apontam na mesma direção. No experimento 1, os pesquisadores investigaram se terminações típicas do feminino (por exemplo, *-ad*, *-íon*) ou masculino (por exemplo, *-ón*) são relevantes no processamento da atribuição de gênero no espanhol na mesma medida em que são as terminações *-a* e *-o*. O teste ANOVA medidas repetidas mostrou que apenas nomes terminados em *-a* e *-o* afetaram a velocidade de resposta dos sujeitos. Esse resultado sugere que o processamento das regularidades fonológicas das terminações *-a* e *-o* é dissociado do processamento de outras terminações, apesar de essas estarem frequentemente associadas ao gênero feminino ou masculino. Portanto, assim como Gollan e Frost (2001), o estudo de Afonso *et al.*(2013) sugere uma dupla rota no processamento do gênero gramatical do espanhol.

Estudos anteriores aos de Gollan e Frost e Afonso *et al.* (2013) para outras línguas, por sua vez, apresentam resultados divergentes. Taft e Meunier (1998) encontraram efeitos de regularidade em uma tarefa de decisão de gênero, mas, também, efeitos de frequência para itens regulares/transparentes quanto ao gênero do francês. Diferentemente do estudo de Gollan e Frost (2001) para o hebraico que contou com estímulos equivalentes em frequência, nos experimentos conduzidos por Taft e Meunier para o francês, a frequência de ocorrência dos itens lexicais selecionados para o experimento foi manipulada. No Experimento 1, os resultados mostraram efeitos de frequência no tempo de reação dos sujeitos para ambas as categorias de formas: regulares e

irregulares¹⁰. Contrariamente à hipótese levantada pelos autores, uma interação entre frequência e regularidade não foi encontrada, sugerindo armazenamento dos itens cuja a terminação era típica de um gênero (regulares) e dos itens cuja terminação não era preditiva de seu gênero (irregulares). Resultados semelhantes haviam sido apresentados anteriormente por Desrochers *et al.* (1989) em um estudo também para o francês. Nesse estudo, efeitos de frequência foram reportados, mas uma interação entre frequência e regularidade não foi encontrada. Essa ausência de interação revela que a frequência influenciou igualmente o tempo de reação dos participantes para ambas as categorias de substantivos: regulares e irregulares. Juntos, os resultados desses estudos sugerem armazenamento de todos os itens, independentemente de sua terminação, isto é, para processar o gênero, os sujeitos acessaram todos os itens na memória.

Embora a frequência pareça exercer um efeito consistente no processamento do gênero, no estudo de Bates *et al.* (1995) para o italiano, efeitos de regularidade foram encontrados, mas efeitos de frequência de ocorrência dos itens não foi estatisticamente relevante em nenhuma das análises empregadas. No entanto, a explicação dos autores para a ausência de efeitos de frequência é relativa ao desenho experimental adotado. Enquanto os experimentos de Gollan e Frost (2001), Taft e Meunier (1998) e Desrochers *et al.* (1989) foram conduzidos visualmente, Bates *et al.* (1995) conduziram um experimento auditivo de decisão de gênero. Segundo os autores, efeitos de frequência na modalidade auditiva são ainda controversos, ao passo que efeitos de frequência são recorrentes na modalidade visual (cf. Bates *et al.* 1995, p. 857).

Em vista dos resultados controversos descritos acima, a questão que levanto no presente estudo é: Se o processamento do gênero é possivelmente governado por um mecanismo responsável pela computação de regras como mostraram Tucker, Lambert, Rigault e Segalowitz (1968), Gollan e Frost (2001) e Afonso *et al.* (2013), como é possível encontrar efeitos de frequência no processamento do gênero de itens regulares quanto ao gênero? O sistema de atribuição de gênero seria então, como hipotetiza o *network model*, um sistema unitário de natureza associativa em que propriedades estruturais mais frequentes prevalecem sobre a propriedades estruturais menos frequentes?

¹⁰ os autores tratam como irregulares todas as formas do francês que não apresentam em sua estrutura ortográfica pistas para reconhecimento de seu gênero.

Neste trabalho, pretendo dar prosseguimento ao debate sobre o processamento do gênero gramatical dos nomes substantivos inanimados. Busco entender se o processamento do gênero de substantivos inanimados do português é governado por uma dupla rota ou se o gênero de itens lexicais é governado por um única sistema de memória em que processos associativos se estabelecem entre traços estruturais compartilhados. Para essa finalidade, conduzi três experimentos distintos. Cada um dos experimentos teve como objetivo observar a influência de dois fatores em diferentes tarefas envolvendo o processamento do gênero gramatical do português: estrutura fonológica dos itens lexicais (principalmente fonemas terminais) e frequência de ocorrência dos itens. No Experimento 1, exploro o processamento da concordância de gênero entre artigos e substantivos inanimados da categoria dos regulares e opacos. Seguindo a tendência da literatura, exploro, no Experimento 2, a influência desses dois fatores em uma tarefa visual de atribuição de gênero a substantivos inanimados de todas as categorias de forma (cf. Cap.2). Por fim, no Experimento 3, usando uma tarefa visual de atribuição de gênero a pseudopalavras, investigo se apenas a terminação ou a estrutura fonológica dos itens lexicais como um todo exerce um impacto no processamento do gênero. Acredito que explorar diferentes aspectos do processamento do gênero gramatical fornecerá um conjunto mais amplo de evidências e, assim, uma interpretação mais precisa dos resultados.

Para analisar o efeito dos diferentes fatores sobre a velocidade de resposta dos sujeitos, os dados resultantes dos Experimentos 1, 2 e 3 foram modelados por um Modelo de Regressão Linear de Efeitos Mistos¹¹ (doravante MEM). Na estrutura do MEM há dois tipos de efeitos: aleatórios (*random effects*) e fixos (*fixed effects*). Os efeitos aleatórios representam uma amostra aleatória de certa população que se deseja avaliar. Os efeitos fixos são as variáveis que hipoteticamente exercem um efeito sobre uma variável dependente (nos experimentos, a velocidade de reposta) e que são submetidas a testes de hipóteses.

Em comparação ao modelo ANOVA medidas repetidas tradicionalmente empregado em experimentos Psicolinguísticos, o MEM apresenta um poder estatístico superior, já que dá conta, ao mesmo tempo, da dependência dos dados entre participante e entre itens.

Uma outra vantagem do MEM reside no fato desse modelo ser mais eficaz (apresentar coeficientes mais precisos) que o teste ANOVA

¹¹ do inglês *Linear Mixed Effects Model*

medidas repetidas para lidar com desenhos experimentais não balanceados (cf. BARR *et al.*, 2013).

A seguir, apresento a metodologia empregada nos Experimentos 1, 2 e 3 e os resultados obtidos pela modelagem dos dados utilizando o MEM.

4.2 Experimento 1

O Experimento 1 contou com uma tarefa de concordância de gênero visualmente apresentada aos sujeitos. O objetivo desse experimento foi testar se o gênero de itens regulares em sua forma quanto ao gênero do português (substantivos femininos terminadas em *-a*, e masculinos terminadas em *-o*) é processado por um mecanismo distinto de itens opacos. Mais especificamente, pretendo verificar se, por um lado, um mecanismo computacional é ativado no processamento do gênero de itens terminados em *-a* e *-o* e, por outro, um sistema de memória está envolvido no processamento do gênero de itens opacos, isto é, itens cujo gênero gramatical não pode ser deduzido de sua estrutura fonológica como as palavras *noite*, *abacaxi* e *lápiz*. Se o processamento do gênero de regularidades fonológicas ativa um mecanismo de natureza computacional, então efeitos de frequência não incidirão sobre itens regulares. Portanto, um efeito da forma bem como uma interação entre forma e frequência do substantivo serão encontrados, sugerindo, respectivamente, que itens opacos e regulares apresentam um processamento distinto e que a frequência de ocorrência dos substantivos afeta o processamento do gênero de itens regulares e itens opacos de maneira distinta.

Se, por outro lado, itens regulares e itens opacos são extraídos da memória para o processamento do gênero, então um efeito de frequência independente (sem interação com o fator forma) será estatisticamente relevante. Será possível encontrar, também, um efeito independente do fator *Forma*, indicando que, mais que computação de regras, o que está em jogo é a força da representação das propriedades fonológicas dos itens, conforme preconiza o *Network Model* (cf. Cap. 3).

4.2.1 Metodologia

Participantes. O Experimento 1 contou com a participação de 19 sujeitos entre 24 e 50 anos de idade (média= 28,7), destros, com visão

normal ou corrigida. Quando questionados pelo experimentador, nenhum dos sujeitos recrutados para o experimento relatou ter desordens neurológicas. Os sujeitos foram recrutados entre estudantes de graduação e pós graduação do Centro de Comunicação e Expressão da Universidade Federal de Santa Catarina e assinaram um termo consentimento de participação livre e esclarecido (Apêndice A).

Materiais. O experimento contou com 56 itens lexicais no total: 32 substantivos e 24 pseudopalavras. Os itens foram selecionados do *Corpus Brasileiro*¹² disponível no sítio da *Linguatca*. Dos 32 substantivos selecionados, 16 itens eram regulares quanto ao gênero (femininos terminados em *-a*, e masculinos terminados em *-o*), e 16 itens opacos quanto ao gênero (sem marcação de gênero).

Metade dos substantivos regulares (8 itens) e metade dos substantivos opacos (8 itens) eram de frequência média a alta (393-1500 por milhão) e a outra metade de baixa frequência (de 0-9 por milhão). No Experimento 1, a distribuição da frequência dos itens foi equiparada entre as categorias: formas regulares frequentes não eram estatisticamente diferentes de formas opacas frequentes ($t = 1.1884$, $df = 11.7$, $p\text{-value} = 0.26$). Semelhantemente, formas regulares infrequentes não eram estatisticamente diferentes de formas opacas infrequentes ($t = 1.1074$, $df = 10.705$, $p\text{-value} = 0.3$). As pseudopalavras foram tratadas como itens distratores (*fillers*) e, por isso, foram descartadas das análises. O conjunto de estímulos utilizado no Experimento 1 bem como a frequência de ocorrência de cada um dos itens lexicais selecionados constam no Apêndice B desta tese.

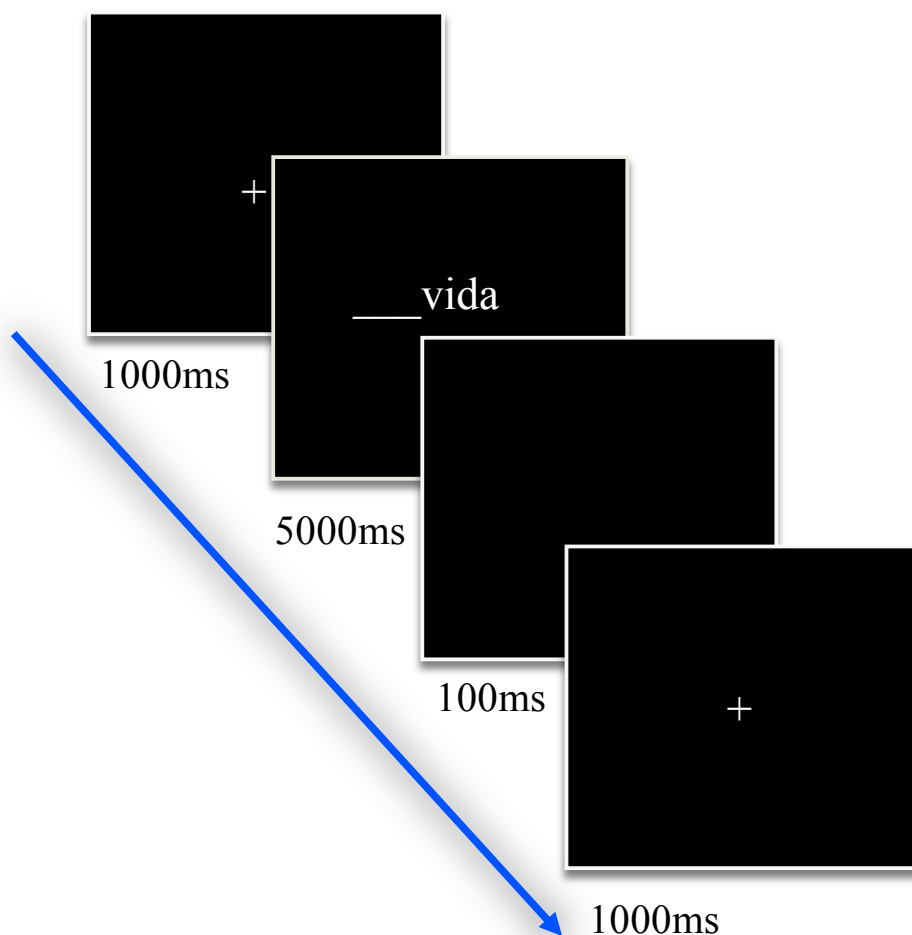
Procedimentos. O Experimento 1 foi programado no software *E-prime 2.0*. Os estímulos foram apresentados visualmente aos sujeitos em caixa baixa, tamanho 18, fonte em cor branca contra um fundo preto de um monitor LG (Flatron L1755S). Os sujeitos participaram da tarefa individualmente na sala do laboratório da Linguagem e Processos Cognitivos da UFSC (*LabLing*). Os procedimentos foram explicados oralmente com a ajuda de instruções na tela do computador. O experimento consistiu em uma tarefa de concordância de gênero. Os sujeitos foram solicitados a pressionar o botão 1 da caixa de respostas

¹² O *Corpus Brasileiro* representa uma coletânea de textos de diversos gêneros do discurso escritos em português brasileiro. O corpus contém mais de 1 bilhão de palavras. O *Corpus Brasileiro* está disponível para consulta no sítio da *Linguatca*: www.linguatca.pt.

para a letra “A” ou botão 5 para a letra “O” os quais correspondiam, respectivamente, aos artigos definidos “O” e “A”.

A apresentação de cada item foi precedida de um sinal de fixação ‘+’ que permanecia na tela por 1000ms. Entre os estímulos, uma tela vazia aparecia por 100ms. Os itens lexicais permaneciam na tela por 5000ms ou até que um dos botões fosse pressionado. A Figura 1, a seguir, ilustra a apresentação dos estímulos.

Figura 1: Apresentação dos Estímulos. Experimento 1.



Abordagem estatística. Os efeitos fixos incluídos no MEM do Experimento 1 foram: *Forma* (Regular/Opaca); *Sexo* do sujeito (Homem/Mulher); *Frequência* (Frequência de ocorrência de cada item no *Corpus Brasileiro* - variável contínua). Como variável controle, o número de *Grafemas* de cada item lexical foi incluído na modelagem (variável contínua). Os efeitos aleatórios foram: *Sujeitos* (19 sujeitos que

participaram do experimento) e *Estímulos* (os 32 substantivos apresentados aos sujeitos).

4.2.2 Resultados

A Tabela 1, abaixo, apresenta a descrição da velocidade de resposta e a Tabela 2 os efeitos encontrados pelo MEM.

Tabela 1: Descrição da velocidade de resposta por categoria de estímulo. Experimento 1. DP (desvio padrão).

Categoria	Min .	Primeir o quartil	Median a	Média	Terceir o quartil	Max.	DP
Substantivos regulares frequentes	540 ms	711 ms	843 ms	890 ms	1000 ms	2115 ms	251.4 ms
Substantivos regulares infrequentes	590 ms	1170 ms	1504 ms	1647 ms	1967 ms	4168 ms	688.8 ms
Substantivos opacos frequentes	567 ms	828 ms	992 ms	1033 ms	1165 ms	2382 ms	304.8 ms
Substantivos opacos infrequentes	615 ms	867 ms	1056 ms	1184 ms	1387 ms	3156 ms	462.4 ms

A descrição dos dados revela que os sujeitos foram mais rápidos, em média, para a concordância com substantivos regulares frequentes que para infrequentes (890ms vs. 1647ms). Da mesma maneira, os participantes foram, em média, mais rápidos para realizar a concordância com substantivos opacos frequentes que substantivos opacos infrequentes (992ms vs. 1056ms). Nota-se, no entanto, que o efeito da frequência foi maior entre itens regulares (frequentes vs. infrequentes) que entre itens opacos (frequentes vs. infrequentes). A seguir, os *Diagramas de Caixa* da Figura 2 ilustram o efeito do fator

Frequência na velocidade de resposta dos participantes para ambas as categorias de forma. A Tabela 2, a seguir, apresenta as variáveis incluídas no modelo que melhor explicam a velocidade de resposta de 19 sujeitos no Experimento 1.

Figura 2: *Diagramas de caixa* da velocidade de resposta por categoria de estímulo. O (opacas), R (regulares).

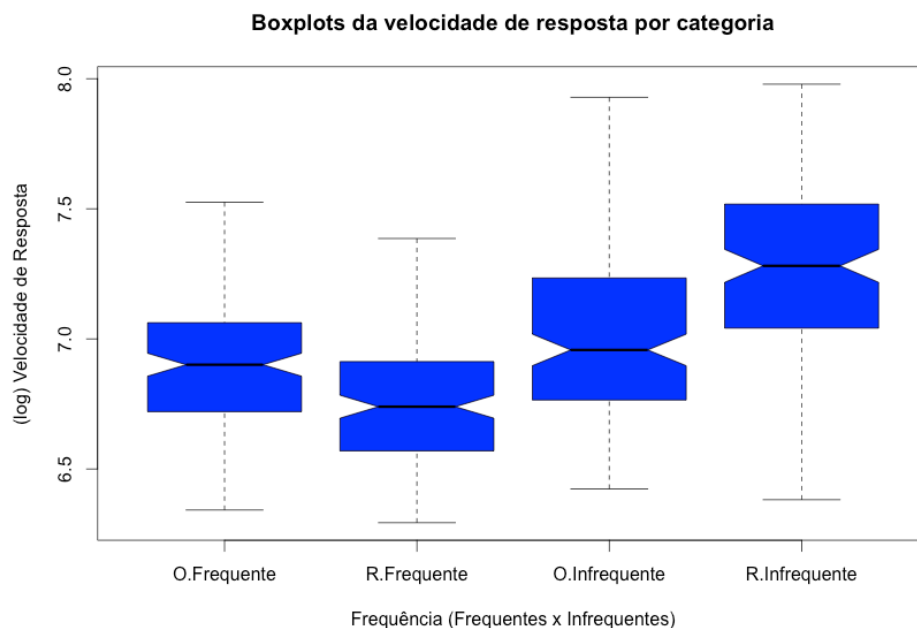


Tabela 2: Variáveis incluídas no modelo que melhor explicam a velocidade de resposta de 19 sujeitos no Experimento 1.

Efeitos fixos	Estimativa	Erro padrão	t-valor	p-valor
intercept	7.038e+00	2.013e-01	34.96	< 0,01
Forma	2.153e-01	7.837e-02	2.75	< 0,05
Forma:FreqSu bs	-3.860e-04	1.180e-04	-3.27	< 0,01
Efeitos Aleatórios	Nome	Variância	Desvio Padrão	Correlação
Estímulos	intercept	0.025492	0.15966	
Sujeito	intercept	0.022938	0.15145	
	Forma	0.001522	0.03901	0.59
Resíduo	0.061074	0.24713		

O tratamento estatístico dos dados foi executado em linguagem *R*. Valores com desvio padrão de 2,5 pontos abaixo e acima da média (*outliers*) foram removidos (2,1% dos dados). Como a distribuição da variável dependente (velocidade de resposta) desviava da normalidade, optei pela transformação logarítmica (log natural) a qual se mostrou mais eficiente na atenuação dos desvios da distribuição que a transformação raiz quadrada (*squareroot transformation*). As análises do Experimento 2 foram realizadas apenas para respostas corretas.

Após a transformação e limpeza do conjunto de dados, quatro diferentes MEM foram construídos, usando o pacote *lme4* (BATES *et al.*, 2014). O primeiro modelo não incluía interação entre as variáveis fixas (*fixed effects*) e os outros cinco modelos incluíram interações entre as variáveis *Forma*, *Frequência* e entre *Forma*, *Frequência* e *Sexo*. A significância das interações incluídas foi verificada por meio do *likelihood ratio test*, pacote *stats* (R DEVELOPMENT CORE TEAM, 2013). A inspeção visual do gráfico de distribuição de resíduos do

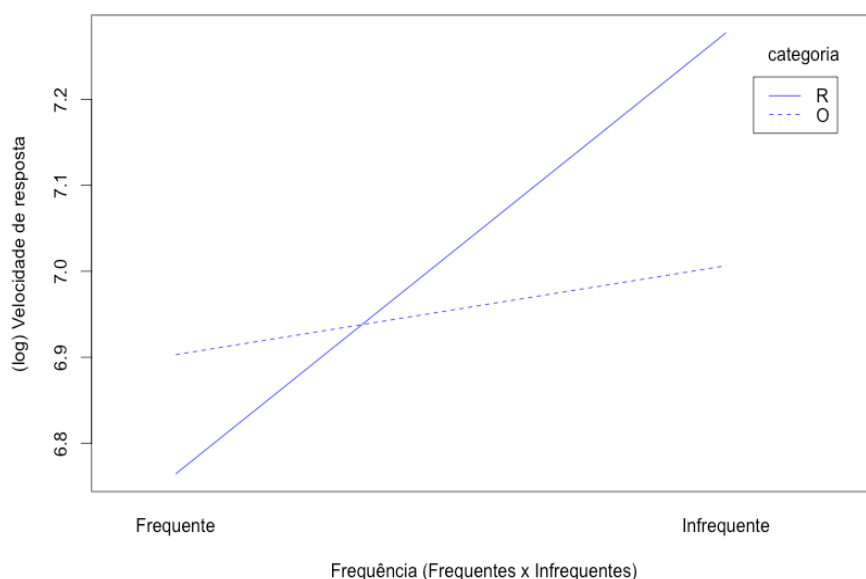
modelo escolhido não revelou anormalidade na distribuição dos resíduos. Os valores de p foram obtidos por meio do teste ANOVA, tipo 3, teste 'F'.

Como se pode observar na Tabela 2 acima, houve um efeito do fator *Forma* ($t=2.75$, $p=0.01$), e uma interação entre os fatores *Forma* e *Frequência* ($t=-3.27$, $p=0.02$). As variáveis *Grafema*, *Sexo* não foram estatisticamente significativas ($p>0.05$). Também não foi significativa uma interação entre *Sexo*, *Forma* e *Frequência* ($p>0.05$).

Em uma análise exploratória dos dados, usei o teste t pareado para investigar o efeito da frequência na variável *Forma*. Os testes t revelaram que formas regulares frequentes são estatisticamente diferentes de formas regulares infrequentes ($t = -14.2435$, $df = 266.859$, $p < 0.01$). Uma diferença foi encontrada, também, entre formas opacas frequentes e infrequentes ($t = -2.9157$, $df = 283.245$, $p\text{-valor} < 0.05$) e entre regulares e opacas frequentes ($t = -4.4469$, $df = 298.923$, $p < 0.01$) bem como entre regulares e opacas infrequentes ($t = 6.8105$, $df = 286.949$, $p < 0.01$).

A Figura 3, a seguir, exibe o efeito da interação entre os fatores *Forma* e *Frequência* dos substantivos.

Figura 3: Gráfico de interação de dois fatores: Forma e Frequência. Experimento 1.



4.2.3 Discussão dos resultados

Os resultados claramente mostram que os fatores *Forma* e *Frequência* exercem um impacto no processamento do gênero gramatical de nomes substantivos inanimados do português. Por outro lado, a interação encontrada entre os fatores *Forma* e *Frequência* mostra que o fator *Frequência* exerce um efeito no processamento do gênero tanto em substantivos regulares quanto em substantivos opacos quanto ao gênero, mas de maneira distinta. Conforme ilustra o gráfico de interação (Figura 3), os sujeitos foram mais rápidos, em média, para itens regulares frequentes que para itens opacos frequentes. O gráfico de interação revela, também, que, quando os fatores *Forma* (regular) e *Frequência* (alta) são combinados, o processamento tende ser mais rápido em comparação ao processamento de itens opacos frequentes. Em acréscimo, é possível observar que os sujeitos foram mais lentos para itens regulares infrequentes que para itens opacos infrequentes.

Uma explicação possível para esse resultado seria a ativação do mecanismo computacional no processamento de itens regulares infrequentes e acesso na memória de itens frequentes em sua forma inteira. Enquanto todos os itens opacos seriam buscados na memória para o processamento do gênero, a ativação do sistema computacional para os itens regulares infrequentes seria um processo mais lento, já que a forma sofreria um processo de decomposição para fins de concordância no sintagma nominal. Alegre e Gordon (1999), por exemplo, mostraram que, abaixo de 6 por milhão, um sistema computacional seria ativado, resultando em um processamento mais lento. Por outro lado, se um sistema de computação de regras é ativado para formas infrequentes, por que efeitos de frequência incidem sobre a categoria de itens regulares? Se, como sugere o estudo de Tucker *et al.* (1968) para o francês usando pseudopalavras, a computação da regra da terminação para a concordância de gênero é independente do acesso lexical, por que na concordância de gênero com formas de terminação regular o itens seriam acessadas na memória (léxico mental)? Como já mencionado, efeitos de frequência na atribuição de gênero a itens regulares quanto ao gênero foram encontrados no estudo de Desrochers (1989) e Taft e Meunier (1998) (cf. Seção 4.1). Vários outros estudos mostraram que itens regulares podem ser memorizados (por exemplo, BAAYENS; DIJKSTRA; SCHREUDER, 1997; ALEGRE; GORDON, 1999). Pinker e Ullman (2002) explicam que é possível que itens altamente regulares sejam armazenados, isto é, processados sem análise

de sua estrutura interna, pelo fato de estarem facilmente disponíveis para acesso na memória dos falantes.

Em vista das evidências de memorização de itens regulares frequentes no Experimento 1, a pergunta que levanto é: Seria, então, o gênero gramatical de itens regulares e opacos processado por um único mecanismo de memória associativa em que prevalecem os esquemas (*schemas*) de associações mais frequentes como sugere o *Network Model* ou por um sistema de processamento dual? (cf. Cap. 3). Em outros termos, o processamento do gênero de formas regulares seria o produto da computação de uma regra ou de um *schema* associativo frequente? Se o processamento do gênero gramatical do português for governado por apenas um único sistema de memória associativa como propõe o *Network Model* e modelos conexionistas, todos os itens estão sujeitos a efeitos de frequência. Além disso, há uma gradação na força da representação das propriedades fonológicas dos itens: o processamento do gênero de formas lexicais que exibem propriedades fonológicas de maior força associativa (propriedades mais frequentes) distingue-se do processamento do gênero de itens lexicais que exibem propriedades fonológicas de menor força associativa (propriedades menos frequentes). Assim sendo, para investigar a hipótese da ativação de um sistema de memória associativa no processamento do gênero gramatical do português, é preciso comparar o processamento de substantivos com terminações frequentemente e infrequentemente associadas aos gêneros feminino e masculino com o objetivo de verificar se há uma gradação na força de representação das propriedades fonológicas do itens e sua associação com as categorias de gênero. Para dar prosseguimento a essa discussão, comparo, no Experimento 2, o processamento de itens lexicais regulares (femininos terminados em *-a*, e masculinos terminados em *-o*) com 3 outras categorias de formas: itens cuja terminação é típica do gênero feminino (terminados em *-agem* e *-ade*), e do gênero masculino (terminados em *-or* e *-ume*), itens cuja terminação contradiz seu gênero (terminados em *-ema* e *-ão*) e itens opacos quanto o gênero (outras terminações). Como visto no Capítulo 2 desta tese, no português, a terminação *-a* está mais frequentemente associada ao gênero feminino, e terminação *-o* mais altamente associada ao gênero masculino que outras terminações, mesmo aquelas que são típicas do gênero feminino, como *-agem* e *-ade*, ou típicas do gênero masculino, tais como: *-or*, *-ema*, *-ume*.

No Experimento 2, será possível observar se, como afirmam os modelos de via dual, somente as regularidades ativam um sistema computacional enquanto as irregularidades e subregularidades seriam

processadas por um sistema de memória associativa. Em acréscimo, será possível observar em que medida as terminações regulares *-a* e *-o* distingue-se da transparência de outras terminações como *-agem*, *-ade*, *-or* e *-ume*.

4.3 Experimento 2

No Experimento 1, um efeito do fator *Forma* foi encontrado, sugerindo uma distinção entre os processamentos de itens opacos e itens regulares quanto ao gênero. A interação entre os fatores *Forma* e *Frequência* encontrada, sugere que, embora diferentemente, a frequência de ocorrência dos itens influenciou o processamento tanto de itens regulares quanto de itens opacos. O Experimento 1 mostrou, também, que o tamanho do efeito da frequência foi maior entre itens regulares que entre itens opacos.

Quando analisados à luz das hipóteses levantadas pelos modelos de via dual e modelos de via única, os resultados obtidos no Experimento 1 não fornecem argumentos suficientes para encaixar o processamento do gênero gramatical do português em um ou outro arcabouço teórico, já que os resultados sugerem, por um lado, armazenamento de formas regulares frequentes e, de outro, a ativação de um mecanismo computacional para as formas regulares infrequentes. Com o objetivo de buscar mais dados para uma interpretação mais precisa dos resultados, proponho, no Experimento 2, incluir duas outras categorias de formas na investigação do processamento do gênero gramatical do português, a saber: itens irregulares quanto ao gênero (substantivos femininos terminados em *-o*, e masculinos terminados em *-a*) e itens cuja terminação são típicas do gênero feminino como *-agem* e *-ade*, e do gênero masculino como *-or*, *-ume* (cf. Cap. 2). Sendo assim, diferentemente do Experimento 1 que comparou o processamento de itens regulares com o processamento de itens opacos, no Experimento 2, serão contrastadas 4 categorias de formas: itens regulares (femininos terminados em *-a*, e masculinos terminados em *-o*) e irregulares (femininos terminados em *-o*, e masculinos terminados em *-a*); itens opacos quanto ao gênero (outras terminações) e, por fim, itens cuja terminação são típicas do gênero feminino e masculino. Farei referência a esta última categoria de itens como *Transparentes*, para diferenciá-los dos itens regulares (cf. Cap. 2).

O objetivo do Experimento 2 é investigar se, assim como preconiza o *Network Model*, o processamento do gênero gramatical do português é governado por um único sistema de caráter associativo em

que, ao invés de decomposição e computação de regras, o que está em jogo é a força da representação das propriedades fonológicas dos itens lexicais, ou se, por outro lado, o processamento do gênero gramatical é comandado por uma dupla rota: um sistema computacional para o processamento do gênero das regularidades e um sistema associativo para o processamento do gênero das irregularidades, incluindo itens subregulares como, por exemplo, o subconjunto de verbos do inglês que apresentam alternância de vogais para formar o passado (ex: *drive-drove*, *write-wrote*, *fly-flew*, *blow-blew*), ou no caso do gênero gramatical do português, itens que apresentam propriedades fonológicas (terminações diferentes de *-a* e *-o*) que podem ser associadas ao gênero feminino (*-agem*, *-ade*) ou gênero masculino (*-or*, *-ume*) (cf. Cap. 3).

As hipóteses que levanto para o Experimento 2, são: Se o gênero gramatical é processado por um único sistema de memória, então diferenças na velocidade de resposta dos sujeitos para o processamento do gênero de cada categoria de forma estarão relacionadas a diferenças na força da representação das propriedades fonológicas (frequência de ocorrência da propriedade estrutural) dos itens (neste caso, terminação) com uma determinada categoria de gênero. Sendo assim, efeitos de regularidade devem ser encontrados para itens regulares e transparentes quanto ao gênero. Além disso, o tamanho do efeito da frequência de ocorrência será similar entre as categorias, portanto, uma interação entre frequência e categoria de forma não será estatisticamente significativa. Se, por outro lado, o processamento do gênero é governado por um sistema dual, então efeitos de regularidade serão encontrados apenas para as formas regulares. Os resultados mostrarão, também, uma interação significativa entre os fatores *Forma* e *Frequência*, sugerindo que o efeito da frequência distingue-se entre as categorias.

4.3.1 Metodologia

Participantes. Para o Experimento 2, foram recrutados 30 sujeitos entre 18 e 48 anos de idade (média= 24 anos), 29 destros e 1 canhoto, com visão normal ou corrigida. Quando questionados, nenhum dos participantes recrutados para o experimento relatou ter desordens neurológicas. Os sujeitos foram recrutados entre estudantes de graduação e pós graduação do Centro de Comunicação e Expressão da Universidade Federal de Santa Catarina. Todos os sujeitos assinaram um consentimento de participação livre e esclarecido (Apêndice A).

Materiais. Para o Experimento 2, 96 substantivos inanimados foram selecionados do *Corpus Brasileiro*. Metade desses substantivos

(48 substantivos) eram regulares quanto ao gênero (femininos terminados em *-a*, masculinos terminados em *-o*). A outra metade foi dividida entre: irregulares (femininos terminados em *-o*, masculinos terminados em *-a*), transparentes (terminações típicas do gênero feminino como *-agem* e *-ade*, e típicas do gênero masculino como *-or* e *-ume*) e opacas quanto ao gênero (outras terminações). A Tabela 3, abaixo, apresenta a quantidade de itens por tipo de terminação.

Tabela 3: Categorias de Formas e quantidade de substantivos por terminação no Experimento 2.

Categori a de Forma	Terminação	Quantidade	Exemplo
Regulares	<i>-a</i>	24	<i>vida</i>
	<i>-o</i>	24	<i>acordo</i>
Transpare ntes	<i>-ade</i>	6	<i>cidade</i>
	<i>-agem</i>	6	<i>imagem</i>
	<i>-or</i>	6	<i>valor</i>
	<i>-ume</i>	6	<i>ciúme</i>
Irregulare s	<i>-ão</i>	6	<i>região</i>
	<i>-ema</i>	6	<i>problema</i>
Opacas	<i>outras terminações</i>	12	<i>nome</i>
Total		96 itens	

Note que, cada terminação pertence a uma categoria de *Forma*. Metade das terminações de cada categoria de forma são substantivos frequentes (frequência média e alta) (165 a 958 por milhão) e a outra metade de substantivos infrequentes (1-13 por milhão). No Experimento 2, a distribuição da frequência dos itens foi equiparada entre as

categorias: formas regulares frequentes não eram estatisticamente diferentes de formas transparentes, irregulares e opacas frequentes ($t = 1.1179$, $df = 42.41$, $p\text{-value} = 0.27$). Semelhantemente, formas regulares infrequentes não eram estatisticamente diferentes de formas transparentes, irregulares e opacas infrequentes ($t = -0.6725$, $df = 42.291$, $p\text{-value} = 0.5$). O conjunto de estímulos utilizado no Experimento 2 bem como a frequência de ocorrência de cada um dos itens lexicais selecionados constam no Apêndice C desta tese.

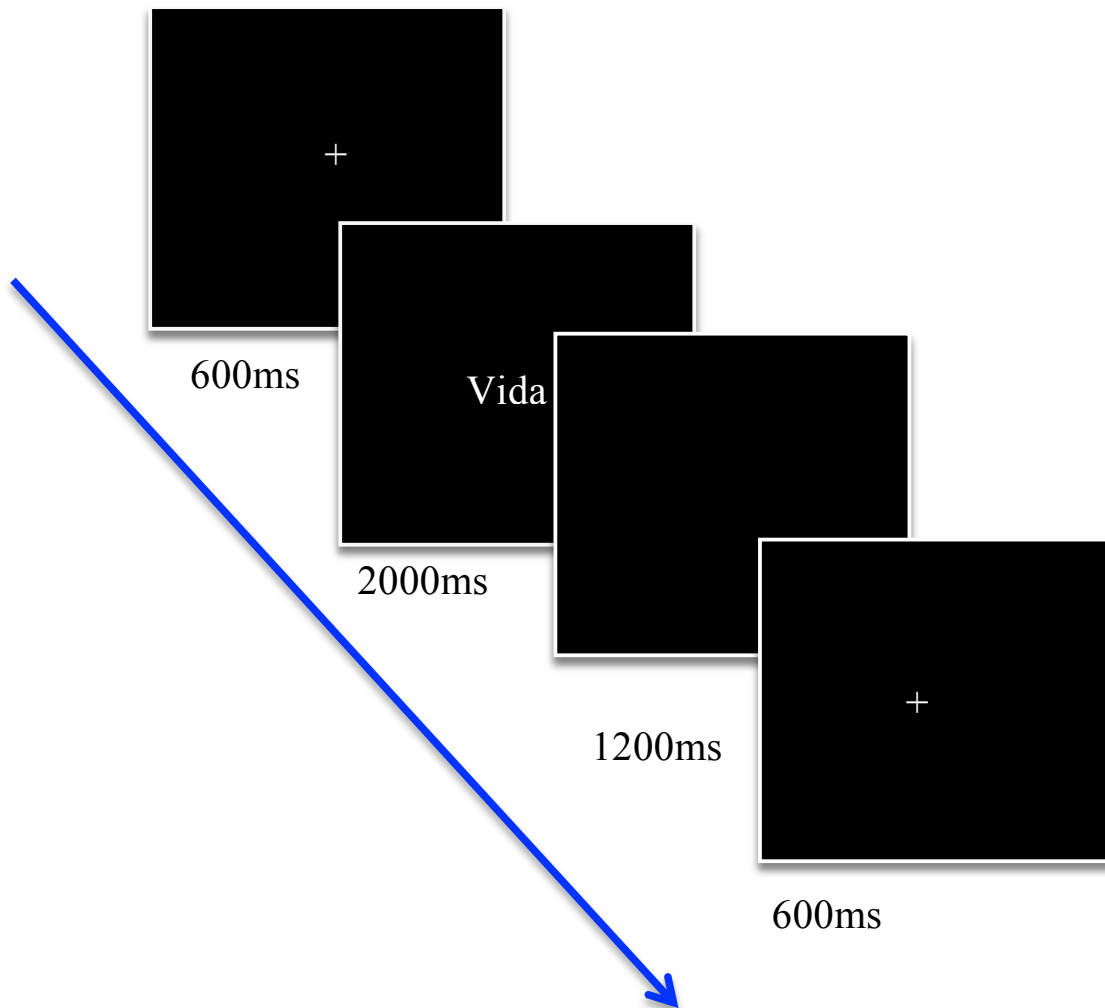
Procedimentos. Seguindo uma tendência na literatura, o desenho experimental adotado no Experimento 2 contou com uma tarefa de atribuição de gênero, isto é, os sujeitos deveriam decidir se a palavra que aparecia na tela do computador era feminina ou masculina.

O experimento foi programado no software *E-prime 2.0*. Os estímulos foram apresentados em caixa baixa, sendo apenas a primeira letra de cada item maiúscula. Os itens foram apresentados em fonte tamanho 18 em cor branca contra um fundo preto de um monitor LG (Flatron L1755S). Os sujeitos participaram da tarefa individualmente na sala do laboratório da Linguagem e Processos Cognitivos da UFSC (*LabLing*). Os estímulos foram apresentados visualmente. Os procedimentos de coleta foram explicados oralmente com a ajuda de instruções na tela do computador.

Os sujeitos foram solicitados a pressionarem a tecla 1 do teclado do computador para o gênero feminino, ou a tecla 3 para o gênero masculino. A apresentação de cada item foi precedida de um sinal de fixação ‘+’ que permanecia na tela por 600ms. Os itens permaneciam na tela por 2000ms ou até que uma das teclas fosse pressionada. A apresentação dos itens foi precedida de uma tela vazia com duração de 1200ms. O desenho experimental foi baseado em Baayens, Dijkstra, Schreuder (1997).

O Experimento 2 foi dividido em 2 blocos, cada bloco contendo 48 itens: 6 itens de cada categoria. Uma pausa para descanso foi introduzida entre os blocos. Os sujeitos foram instruídos a descansarem pelo tempo que achassem necessário. Os itens de cada bloco e a apresentação dos blocos foram randomizados entre os 30 sujeitos. A apresentação do Experimento 2 foi alternada com a apresentação do Experimento 3 para evitar efeitos de ordem. A Figura 4, a seguir, ilustra a apresentação dos estímulos.

Figura 4: Apresentação dos Estímulos. Experimento 2.



Abordagem estatística. Assim como no Experimento 1, os dados foram modelados por um Modelo de Regressão Linear de Efeitos Mistos. Para a modelagem dos dados do Experimento 2, os efeitos aleatórios foram: *Sujeitos* (30 sujeitos testados) e os *Estímulos* (96 substantivos apresentados na tela para atribuição de gênero). Os efeitos fixos incluídos no modelo foram: *Forma* (formas regulares, irregulares, opacos e transparentes quanto ao gênero), *Frequência* (Frequência de ocorrência de lema dos substantivos no *Corpus Brasileiro* - variável contínua) e *Sexo* dos sujeito (homem/mulher). Como variáveis controle, foram incluídas o número de *Grafemas* dos itens apresentados na tela (variável contínua), e a *Ordem* de apresentação da tarefa (Primeira/Segunda).

4.3.2 Resultados

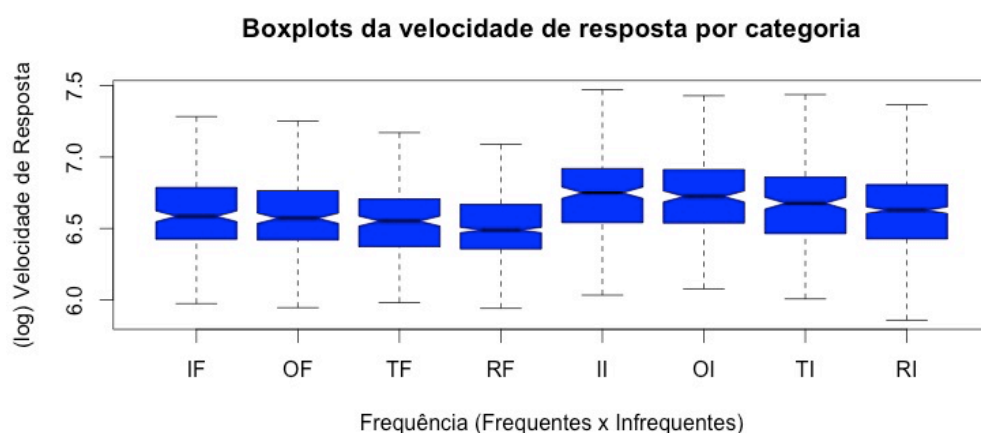
A Tabela 4, abaixo, exhibe a descrição dos dados por categoria de substantivo. A Tabela 5, a seguir, exhibe os efeitos dos fatores estatisticamente relevantes no Experimento 2.

Tabela 4: Descrição da velocidade de resposta de 30 sujeitos na atribuição de gênero.

Categoria	Min.	Primeiro o quartil	Medi ana	Média	Terceiro quartil	Max.	DP
Substantivos regulares frequentes	349 ms	575 ms	657 ms	696 ms	787 ms	1678 ms	186 ms
Substantivos regulares infrequentes	303 ms	617 ms	755 ms	787 ms	904 ms	1816 ms	240 ms
Substantivos irregulares frequentes	392 ms	616 ms	723 ms	765 ms	881 ms	1739 ms	221 ms
Substantivos irregulares infrequentes	348 ms	691 ms	853 ms	877 ms	1009 ms	1848 ms	256 ms
Substantivos opacos frequentes	381 ms	612 ms	713 ms	754 ms	865 ms	1590 ms	223 ms
Substantivos opacos infrequentes	343 ms	693 ms	831 ms	868 ms	1004 ms	1922 ms	282 ms
Substantivos transparentes frequentes	395 ms	584 ms	699 ms	718 ms	817 ms	1621 ms	186 ms
Substantivos transparentes infrequentes	405 ms	643 ms	793 ms	830 ms	949 ms	1964 ms	254 ms

A descrição dos dados revela que, entre todas as categorias, os sujeitos foram mais rápidos, em média, para atribuir gênero aos substantivos regulares frequentes e menos velozes para atribuir gênero a substantivos irregulares. Se compararmos a velocidade de resposta dos sujeitos para atribuir gênero a substantivos regulares e substantivos transparentes, é possível observar a proximidade das médias tanto no que se refere a itens frequentes (696ms vs. 718ms) quanto a itens infrequentes (787ms vs. 830ms) com uma pequena vantagem para os substantivos regulares. Da mesma forma, é possível observar que a média da velocidade de respostas dos sujeitos para atribuir gênero a substantivos opacos aproxima-se da média da velocidade de resposta na atribuição de gênero a substantivos irregulares tanto no que se refere a itens frequentes (713ms vs. 723ms) quanto a itens infrequentes (868ms vs. 877ms) com uma vantagem para os itens opacos. Assim, se agrupadas em uma escala crescente de velocidade de resposta por categoria, temos a seguinte ordenação: substantivos regulares, substantivos transparentes, substantivos opacos e substantivos irregulares. A seguir, os *Diagramas de Caixa* da Figura 5 ilustram essa gradação:

Figura 5: *Diagramas de Caixa* da velocidade de resposta de 30 sujeitos em uma tarefa de atribuição de gênero a 8 categorias de Formas: Irregulares Frequentes (IF) e Infrequentes (II); Regulares Frequentes (RF) e Infrequentes (RI); Transparentes Frequentes (TF) e Infrequentes (TI); Opacas Frequentes (OF) e Infrequentes (OI).



Assim como no Experimento 1, o tratamento estatístico dos dados foi executado em linguagem *R*. Valores com desvio padrão de 2,5 pontos abaixo e acima da média (*outliers*) foram removidos (1,8% dos dados). Como a distribuição da variável dependente (velocidade de resposta) desviava da normalidade, optei pela transformação logarítmica (log natural) a qual se mostrou mais eficaz que a transformação raiz quadrada (*squareroot transformation*) na atenuação do desvio da distribuição. As análises do Experimento 2 foram realizadas apenas para respostas corretas.

Após a transformação e limpeza do conjunto de dados, quatro diferentes MEM foram construídos, usando o pacote *lme4* (BATES *et al.*, 2014). O primeiro modelo não incluía interação entre as variáveis fixas (*fixed effects*) e os outros cinco modelos incluíram interações entre as variáveis *Forma* e *Frequência* e entre *Forma* e *Frequência* e *Sexo*. A significância das interações incluídas foi verificada por meio do *likelihood ratio test*, pacote *stats* (R DEVELOPMENT CORE TEAM, 2013). A inspeção visual do gráfico de distribuição de resíduos do modelo escolhido não revelou anormalidade na distribuição dos resíduos. Os valores de *p* foram obtidos pelo teste ANOVA, tipo 3, teste 'F'.

A Tabela 5, abaixo, exhibe o modelo que melhor explica a velocidade de resposta dos sujeitos para o Experimento 2.

Tabela 5: Variáveis incluídas no modelo que melhor explicam a velocidade de resposta de 30 sujeitos no Experimento 2.

Efeitos fixos	Estimativa	Erro padrão	t-valor	p-valor
<i>intercept</i>	6.734e+00	6.065e-02	111.02	< 0,01
Forma (Transparente)	-6.613e-02	3.256e-02	-2.03	<0,05
Forma (Regular)	-1.177e-01	2.746e-02	-4.28	<0,01
Frequência	-1.485e-04	5.610e-05	-2.65	<0,01
Efeitos Aleatórios	Nome	Variância	Desvio Padrão	

Sujeitos	intercept	0.002942	0.05424	
Estímulos	intercept	0.026868	0.16391	
	Resíduo	0.031950	0.17874	

Como se pode observar na Tabela 5, houve um efeito de *Forma* para as categorias dos substantivos regulares ($t=-4,28$, $p<0,01$) e categoria dos substantivos transparentes ($t=-2,03$, $p<0,01$). Um efeito independente da frequência emergiu como significativo ($t=-2,65$, $p<0,01$). A interação entre os fatores *Frequência* e *Forma* não foi estatisticamente relevante ($p>.1$), nem mesmo uma interação entre os fatores *Sexo*, *Frequência* e *Forma* ($p>0,1$). Também não foram estatisticamente relevantes as variáveis controle *Ordem* e número de *Grafemas*.

4.3.3 Discussão dos resultados

Na descrição dos dados do Experimento 2, é possível observar que a frequência de ocorrência dos substantivos influenciou o processamento da atribuição de gênero independentemente da forma do item. Itens altamente frequentes foram processados mais rapidamente que itens infrequentes. É possível observar, ainda, que os fatores transparência e regularidade das formas fornece uma vantagem sobre o processamento do gênero. A vantagem do fator transparência fica evidente ao comparar as distribuições da velocidade de resposta entre as categorias: os sujeitos foram mais velozes, em média, na atribuição de gênero aos substantivos regulares frequentes (657ms) e infrequentes (755ms) e transparentes frequentes (699ms) e infrequentes (793) que na atribuição de gênero aos substantivos opacos frequentes (713ms) e infrequentes (831ms) e irregulares frequentes (723ms) e infrequentes (853ms).

A inspeção visual da Figura 5 revela que há um ranqueamento na distribuição da velocidade de resposta entre categorias de forma tanto no grupo de itens frequentes quanto no grupo de itens infrequentes: os sujeitos foram mais rápidos na atribuição de gênero, em primeiro lugar, aos substantivos regulares, em segundo lugar, aos substantivos transparentes, em terceiro lugar, aos substantivos opacos e, por último,

aos substantivos irregulares. Esse ranqueamento evidencia que, quanto mais informação sobre o gênero um substantivo apresenta em sua forma, mais veloz é o seu processamento. Se, no entanto, a terminação não fornece informação sobre seu gênero (itens opacos) ou se a terminação do substantivo contradiz a expectativa de seu gênero (itens irregulares), o processamento é dificultado.

O MEM revelou um efeito individual do fator *Frequência*, sugerindo que todas as categorias de formas foram acessadas na memória. A ausência de interação entre os fatores *Forma* e *Frequência* indica que o tamanho do efeito da frequência foi similar entre as categorias de forma. Mais importante, o MEM revelou um efeito tanto para as formas regulares quanto para as formas transparentes. Se, conforme preconiza os modelos de via dual, regularidades fossem processadas por um mecanismo distinto das irregularidades e subregularidades, referenciadas aqui como formas transparentes, então apenas um efeito da categoria de substantivos regulares seria observado. Além disso, uma interação entre forma regular e frequência seria também revelada, sugerindo que o processamento das formas regulares distingue-se do processamento das demais formas. Os dados obtidos no Experimento 2 indicam, no entanto, um cenário oposto. O efeito individual da frequência sugere que todas as formas foram acessadas na memória. Portanto, as diferenças na velocidade de resposta dos sujeitos para atribuir gênero às formas não parece estar relacionada a mecanismos distintos de processamento, mas a diferenças na força de representação das propriedades fonológicas dos itens.

Com já mencionado, os resultados do Experimento 2 mostram que os sujeitos foram mais rápidos para atribuir gênero a substantivos regulares quanto ao gênero que às outras categorias de formas. Portanto, em relação à questão: Em que medida as terminações regulares *-a* e *-o* distingue-se da transparência de outras terminações como *-agem*, *-ade*, *-or* e *-ume*? Os dados aqui obtidos indicam que as terminações *-a* e *-o* apresentam uma força associativa maior que as demais, provavelmente em virtude de sua maior frequência de ocorrência na língua portuguesa. A força associativa de *-a* e *-o* já havia sido demonstrada no estudo conduzido por Afonso *et al.* (2013) para o espanhol (cf. Seção 4.1). Embora itens com outras terminações típicas do gênero feminino e masculino no espanhol estivessem presentes no experimento, apenas *-a* e *-o* afetaram a velocidade de resposta dos sujeitos em uma tarefa de atribuição de gênero. Em acréscimo, a força dessa associação, como visto no Capítulo 2, pode estar ligada, também, ao fato de terminações

em *-a* para o gênero feminino, e terminações em *-o* para o gênero masculino serem, por vezes, morfemas derivacionais distintivos de sexo (*gat-o*, *gat-a* - *cantor*, *cantor-a*), tanto no português quanto no espanhol, quando o substantivo designa um ser animado [+humano]. Outros fatores que possivelmente fortalecem a associação do gênero feminino com a terminação *-a* e do gênero masculino com a terminação *-o* no português são: *a* e *o* são artigos especificadores de gênero. As terminações *-a* e *-o* atuam, também, como sufixos marcadores de gênero em adjetivos (*bonita*, *bonito*) e pronomes do caso reto e caso oblíquo (*o*, *a*, *-los*, *-las*). Afonso *et al.* (2013) argumentam que a frequência de *-a* e *-o* podem ter induzido a uma espécie de ‘gramaticalização’ dessas terminações. No entanto, os resultados do Experimento 2 indicam que, ao invés de gramaticalização, o que parece atuar é a força associativa da terminação *-a* com o gênero feminino, e *-o* com o gênero masculino conforme pressupõe Bybee (1995).

No que se refere à ausência de uma interação entre as variáveis *Sexo* e *Forma* bem como a ausência de interação entre as variáveis *Sexo*, *Forma* e *Frequência*, os dados sugerem que homens e mulheres apresentaram um comportamento semelhante no processamento do gênero gramatical do português. Novamente, esse resultado vai de encontro aos pressupostos dos modelos de via dual, especialmente, vai de encontro aos pressupostos do *Modelo Declarativo/Procedural* (ULLMAN, 1991) que prevê diferenças no processamento da linguagem entre os sexos (cf. Cap. 3).

Em resumo, o Experimento 2 não apresenta evidências que comprovem as premissas assumidas pelos modelos de via dual. Os dados sugerem, no entanto, que o processamento do gênero no português é governado por um único sistema de memória de natureza associativa, já que o experimento mostra que a frequência de ocorrência dos itens influenciou o processamento da atribuição de gênero independentemente da forma do substantivo.

4. 4 Experimento 3

De acordo com o *Network Model* (BYBEE, 1985, 1995), a força da representação das propriedades fonológicas de um item é proveniente de sua frequência de ocorrência (cf. Cap 3). Assim sendo, se uma propriedade fonológica é mais frequente que outra, então a sua força associativa é maior. Nesse sentido, Bybee (1985, 1995) assume que há uma relação entre a força associativa de uma estrutura e seu

processamento: estruturas com maior força associativa são mais facilmente acessadas e, por isso, exibem menor tempo de reação. A visão dual de processamento, em contrapartida, restringe o processamento de tipo associativo às irregularidades e subregularidades. Para os pesquisadores dessa vertente teórica, diferentemente de formas irregulares, em tarefas de compreensão, formas regulares são decompostas e processadas *online*.

Os resultados do Experimento 1 e Experimento 2 sugerem que, independentemente de sua forma, os substantivos foram acessados na memória (léxico mental) dos sujeitos para o processamento de seu gênero. É possível, no entanto, que o processamento do gênero seja um processo realizado em duas etapas: a primeira etapa seria o acesso lexical dos itens na memória (fato que explica efeitos de frequência para todas as formas) e, em uma segunda etapa, a atribuição de gênero aos substantivos seria realizada. Essa hipótese foi levantada primeiramente por Desrochers *et al.* (1989) a qual foi considerada por estudos posteriores como Taft e Meunier (1998) e Bates *et al.* (1995) e Gollan e Frost (2001). Se considerarmos essa hipótese, resta saber se, nessa segunda fase, o que ocorre é um processo associativo para a classificação dos substantivos em categorias de gênero, ou se dois mecanismos são ativados. A pergunta que levanto é: Após acessar os itens no léxico mental, o efeito da força associativa de uma estrutura fonológica atua apenas em itens irregulares, opacos e transparentes ou essa força atua também no processamento das formas regulares quanto ao gênero? No Experimento 3, apresentado a seguir, examino essa questão.

Uma maneira que encontrei de verificar se a força associativa no processamento do gênero se estende às formas regulares quanto ao gênero do português foi bloquear o acesso lexical dos sujeitos por meio de pseudopalavras. Ao bloquear o acesso lexical, foi possível avaliar de forma mais precisa quais são os fatores relevantes no processamento da atribuição de gênero na segunda etapa: apenas a terminação ou a estrutura do item lexical como um todo. No Experimento 3, o processamento da atribuição de gênero a pseudopalavras de maior força associativa e processamento da atribuição de gênero a pseudopalavras de menor força associativa foram comparados.

Tomemos como exemplo as pseudopalavras *vismeda* e *bamília*. Embora a pseudopalavra *vismeda* seja uma estrutura fonotaticamente possível no português, é difícil estabelecer uma associação entre a

estrutura¹³ fônica desse item com outras já pré-existentes no léxico da língua, pelo simples fato de¹⁴ não haver estruturas semelhantes, como pares mínimos ou formas derivadas, para essa estrutura no léxico do português. Em contrapartida, a pseudopalavra *bamília* parece estar em conexão com a palavra *família*, visto que ambos os itens diferem entre si em relação a apenas uma unidade fônica, compartilhando, portanto, traços estruturais. Além disso, a palavra *família* apresenta outros vizinhos ortográficos como as palavras *familiar* e *familiaridade*. Assim sendo, é possível assumir, com base em Bybbe (1995), que a pseudopalavra *vismeda* apresenta uma força associativa menor que a pseudopalavra *bamília*. Em uma tarefa de atribuição de gênero ao item *vismeda*, o falante só teria a terminação *-a* como pista. Por outro lado, em uma tarefa de atribuição de gênero ao item *bamília*, é possível que, ao invés de decompor a forma (*bamili+a*) para atribuir-lhe gênero, o falante leve em consideração a estrutura fônica do item como um todo, estabelecendo, portanto, uma associação entre a estrutura desse item com outras pré-existentes em seu léxico mental, isto é, leve em consideração os traços compartilhados entre os itens. Conforme preconiza o *Network Model*, quanto mais traços compartilhados entre os itens, maior a força associativa, consequentemente, menor o tempo de reação e maior sua acurácia (cf. Cap. 3).

No Experimento 3, lido com dois grupos de pseudopalavras: as fáceis e as difíceis. Considerei como fáceis as pseudopalavras que simulam um par mínimo com uma estrutura pré-existente. As pseudopalavras difíceis foram construídas combinando aleatoriamente sílabas fonotaticamente possíveis do português (CV e CVC). Embora essas construções não firam as regras fonotáticas do português, esses itens não estão em conexão com itens pré-existentes no português, portanto, é possível assumir que apresentam menor força associativa que as pseudopalavras fáceis. Esses dois grupos de pseudopalavras foram apresentados aos sujeitos para que eles atribuíssem gênero às formas. A adoção desse desenho experimental foi baseada no seguinte raciocínio: Se os falantes ativam um mecanismo computacional para atribuição de gênero, então um efeito do grau de dificuldade da pseudopalavra não

¹³ Usarei de agora em diante o termo fônico(a) em substituição ao termo fonológico(a) visto que farei referência a itens inexistentes (pseudopalavras) no léxico do português e, portanto, no léxico mental dos sujeitos.

¹⁴ Em uma busca no dicionário online priberam.pt, o item *vismeda* não foi encontrado nem mesmo vizinhos ortográficos ao item.

será observado no processamento de estruturas regulares (itens terminados em *-a* e *-o*). Se, por outro lado, o processamento do gênero gramatical ativa um mecanismo associativo para todas as formas, então um efeito do grau de dificuldade da pseudopalavra será observado no processamento do gênero de todos os itens apresentados aos sujeitos.

Vários estudos (GOLLAN; FROST, 2001, TAFT, 1994, HOHFELD, 2006) sugerem que pseudopalavras induzem a ativação de um mecanismo computacional. Esses estudos mostram que o tempo de reação a pseudopalavras é maior que a palavras. Baayens, Dijkstra e Schreuder (1997) explicam que a velocidade de resposta a pseudopalavras é mais lenta que a palavras reais em virtude de essas apresentarem um processamento de tipo decomposicional. Assim, será possível observar, com base no tempo de reação a pseudopalavras fáceis e difíceis, se a decomposição se limita a apenas pseudopalavras difíceis (que apresentam apenas a terminação como informação para a atribuição de gênero) ou se esse processo se estende para o processamento das pseudopalavras fáceis (que se distinguem de palavras reais do português em apenas um fonema), principalmente no que se refere aos itens transparentes e regulares quanto ao seu gênero gramatical.

Dado que os resultados obtidos nos Experimentos 1 e Experimento 2 apresentam mais evidências para a hipótese de um mecanismo de tipo associativo para o processamento do gênero dos substantivos inanimados do português, espero encontrar um efeito do fator *Dificuldade* para todos os itens apresentados aos sujeitos independentemente de sua terminação: os participantes serão mais velozes na atribuição de gênero a pseudopalavras fáceis que na atribuição de gênero a pseudopalavras difíceis. Em outros termos, o efeito do fator *Dificuldade* da pseudopalavra será observado para todas as categorias de formas apresentadas aos sujeitos. Esta hipótese está fundamentada em Bybee (1995) que prevê um tempo de reação menor a itens que estão em conexão que a itens cuja rede conexões é fraca (cf. Cap. 3).

Para o Experimento 3, as porcentagens de classificação das pseudopalavras em categorias de gênero serão calculadas. Portanto, espero encontrar um efeito do fator *Dificuldade* da pseudopalavra também nas porcentagens de classificação em categorias de gênero. De forma mais específica, espero que a porcentagem de classificação de pseudopalavras com terminações típicas do feminino (*-a*, *-agem*, *-iade*) seja maior para o gênero feminino, e com terminações típicas do masculino (*-o*, *-or*, *-ume*) seja maior para o gênero masculino. Acredito,

também, que essas porcentagens serão ainda maiores para o grupo de pseudopalavras fáceis.

Em relação ao processamento da estrutura fônica da pseudopalavra, minhas hipóteses são: os sujeitos serão mais velozes, em primeiro lugar, para atribuir gênero a itens terminados em *-a* e *-o* (fáceis e difíceis) que atribuir gênero a itens com outras terminações típicas de uma categoria de gênero (*-agem*, *-dade*, *-or* e *-ume*). Os sujeitos serão mais lentos para atribuir gênero a pseudopalavras terminadas em *-ema* e *-ão* uma vez que essas terminações podem ser associadas tanto ao gênero feminino quanto ao gênero masculino. A seguir, descrevo o desenho experimental adotado no Experimento 3.

4.4.1 Metodologia

Participantes. Participaram do Experimento 3 os mesmos 30 sujeitos que participaram do Experimento 2. Dois participantes foram desconsiderados das análises por terem reportado vários erros durante a execução da tarefa. Portanto, os análises foram baseadas nos dados de 28 sujeitos (27 destros e 1 canhoto).

Materiais. O Experimento 3 contou com 48 pseudopalavras. Esses itens foram divididos em 3 grupos: Pseudopalavras com terminações regulares (*-a* e *-o*), transparentes (*-agem*, *-idade*, *-ume*, *-or*) e irregulares quanto ao gênero (*-ão* e *-ema*). A categoria das regulares continha 12 itens, a categoria das transparentes 24 itens (12 femininos e 12 masculinos) e a das irregulares 12 itens. Esses itens foram divididos em 2 grupos: Pseudopalavras fáceis e Pseudopalavras difíceis.

Conforme mencionado anteriormente, considerei como pseudopalavras fáceis, os itens cuja estrutura fônica forma um par mínimo hipotético com uma palavra existente no português. Por exemplo, (*Família/Bamília*). Considerei como pseudopalavras difíceis, os itens que combinavam aleatoriamente sílabas fonologicamente possíveis no português (por exemplo, *cetoda*). Esses itens foram construídos com diversas terminações: terminações regulares quanto ao gênero (*-a* e *-o*), terminações transparentes, isto é, típicas do gênero feminino e do gênero masculino (*-agem*, *-ade*, *-or*, *-ume*), e itens com terminações ambíguas, ou seja, aquelas que podem ser associadas tanto a gênero feminino quanto ao gênero masculino, são elas: *-ema* e *-ão* (cf. Cap. 2). O conjunto de estímulos utilizado no Experimento 3 consta no Apêndice D desta tese.

A Tabela 6, abaixo, exhibe as categorias de itens e quantidade de itens por categoria.

Tabela 6: Categorias de formas e quantidade de pseudopalavras incluídas no Experimento 3.

Categorias de Formas	Terminação	Fáceis	Díficeis
Terminações Regulares	<i>-a</i>	3	3
	<i>-o</i>	3	3
Terminações Transparentes	<i>-agem</i>	3	3
	<i>-ade</i>	3	3
	<i>-ume</i>	4	2
	<i>-or</i>	3	3
Terminações Irregulares	<i>-ema</i>	3	3
	<i>-ão</i>	3	3
Total de estímulos	8 tipos de terminações	24	24

Procedimentos. O desenho experimental e os procedimentos de coleta do Experimento 3 foram idênticos ao do Experimento 2 (cf. Seção 4.3.1). As 48 pseudopalavras foram apresentadas em apenas 1 bloco. A ordem de apresentação do Experimento 3 foi alternada com a apresentação do Experimento 2. Os estímulos do Experimento 3 foram randomizados entre os participantes para evitar efeitos de ordem.

Abordagem estatística. Assim como no Experimento 1 e Experimento 2, os dados foram modelados por um Modelo de Regressão Linear de Efeitos Mistos. Para a modelagem dos dados do Experimento 3, os efeitos aleatórios foram: *Sujeitos* (28 sujeitos testados) e os *Estímulos* (48 pseudopalavras apresentadas na tela para atribuição de gênero). Os efeitos fixos incluídos no modelo foram: categorias de *Formas* (itens de formas regulares, irregulares, opacos, transparentes), *Terminação* da pseudopalavra (*-a, -o, -agem, -ade, -or, -*

ume, *-ema* e *-ão*), grau de *Dificuldade* da pseudopalavra (fácil/difícil) e *Sexo* dos sujeitos (homem/mulher). Como variável controle, foram incluídos o número de *Grafemas* dos itens apresentados na tela (variável contínua), e a *Ordem* de apresentação da tarefa (Primeira/Segunda).

4.4.2 Resultados

A Tabela 7, abaixo, exhibe as porcentagens de classificação das pseudopalavras em categorias de gênero (feminino vs. masculino).

Tabela 7: Porcentagens de classificação das pseudopalavras em categorias de gênero: Fáceis x Difíceis.

Terminação da pseudopalavra	Difíceis		Fáceis	
	Gênero feminino	Gênero masculino	Gênero feminino	Gênero masculino
<i>-a</i>	82,1%	15,5%	90,5%	9,5%
<i>-o</i>	21,4%	78,6%	1,2%	98,8%
<i>-agem</i>	84,5%	15,5%	84,5%	15,5%
<i>-ade</i>	70,3%	28,6%	77,4%	22,6%
<i>-ume</i>	37,5%	62,5%	21,42%	77,7%
<i>-or</i>	7,1%	92,9%	22,6%	94,4%
<i>-ema</i>	71,4%	28,6%	61,9%	38,1%
<i>-ão</i>	9,5%	90,5%	73,8%	26,2%

Como esperado, os sujeitos atribuíram o gênero feminino à maioria dos itens terminados em *-a* (82,1% aos itens difíceis e 90,5% aos itens fáceis), e o gênero masculino aos terminados em *-o* (78,6% aos itens difíceis e 98,8% aos itens fáceis). No grupo das pseudopalavras com terminações transparentes (típicas do gênero feminino e do gênero masculino), os itens com terminações tipicamente femininas foram classificadas, em sua maioria, como femininas: *-agem* (84,5% para fáceis e difíceis) e *-ade* (70,3% difíceis e 77,4% fáceis), e os itens com terminações tipicamente masculinas foram classificados, em sua maioria, como do gênero masculino: *-ume* (62,5% difíceis e 77,7% fáceis), *-or* (92,9% difíceis e 94,4% fáceis). Nota-se, no entanto, que as

terminações que apresentam a vogal átona final *-e* desse mesmo grupo (*-ade* e *-ume*) receberam a menor porcentagem de acertos quando comparadas às outras terminações da mesma categoria, apesar de serem típicas do gênero feminino e masculino respectivamente. Para minha surpresa, os itens terminados em *-or* receberam maior porcentagem de classificação como masculinas que os terminados em *-o* no grupo das difíceis (*-o* com 78,6% vs. 92,9% para *-or*), sugerindo que a terminação *-or* é mais transparente que *-o*.

Ao compararmos as porcentagens de classificação entre as pseudopalavras fáceis e difíceis, é possível observar, assim como no tempo de reação, um efeito da estrutura dos itens. As porcentagens de classificação das pseudopalavras nas categorias aumentam quando as pseudopalavras oferecem mais que a terminação como pista. Por exemplo, os itens terminados em *-a* foram classificados como femininos em 82,1% das possibilidades de resposta no grupo das difíceis, mas essa porcentagem aumenta para 90,5% para a classificação dos itens no grupo das fáceis. Esse aumento pode ser observado, também, na classificação dos itens terminados *-o* como masculinos (78,6% vs. 98,8%), *-ade* como femininos (70,3% vs. 77,4%) e *-ume* como masculinos (62,5% vs. 77,7%). Em relação às pseudopalavras terminadas em *-agem*, não houve um aumento da porcentagem entre fáceis e difíceis. Interessante notar que, as pseudopalavras difíceis terminadas em *-ema* foram, em sua maioria, classificadas como femininas. Essa porcentagem diminui quando a pseudopalavra pertence ao grupo das fáceis que simulam pares mínimos com palavras masculinas, como (*cinema/cilema*). Esse mesmo efeito pode ser observado numa proporção ainda maior entre as pseudopalavras terminadas em *-ão*: no grupo das difíceis, os itens terminados em *-ão* foram, em sua maioria, classificados como masculinos (90,5%), mas no grupo das fáceis a classificação em masculinas diminui, aumentando para 73, 8% a classificação como femininas. Dado que as pseudopalavras fáceis terminadas em *-ão* são pares mínimos hipotéticos de palavras femininas do português, esse resultado sugere, novamente, o efeito da força associativa da forma como um todo. Assim como preconiza o *Network Model*, quanto mais traços compartilhados entre os itens, menor o seu tempo de reação e maior sua acurácia de resposta (BYBEE, 1995, p.434).

No que se refere à velocidade de resposta dos sujeitos na atribuição de gênero, a Tabela 8, a seguir, apresenta a média, mediana e desvio padrão dos dados obtidos por categoria e por tipo de terminação.

Tabela 8: Descrição da velocidade de resposta de 28 sujeitos na atribuição de gênero a pseudopalavras.

Terminação da pseudopalavra	Min.	Média	Mediana	Max.	DP
-a	343ms	934ms	912ms	1847ms	304ms
-o	321ms	908ms	861ms	1887ms	296ms
-agem	397ms	978ms	920ms	1963ms	304ms
-ade	319ms	1025ms	974ms	1974ms	350ms
-ume	347ms	1016ms	995ms	1956ms	320ms
-or	391ms	939ms	910ms	1946ms	295ms
-ema	366ms	953ms	930ms	1941ms	295ms
-ão	380ms	1022ms	953ms	1992ms	331ms

Como se pode observar na Tabela 8 acima, os sujeitos foram mais rápidos, em média, para atribuir gênero a pseudopalavras com terminações em *-a* e *-o* (934ms e 908ms) que para outras terminações, em especial, para atribuir gênero a itens terminados em *-o*. Entre o grupo das terminações transparentes, é possível observar um efeito facilitador da terminação *-or* (939ms) na atribuição de gênero em relação às outras terminações. Nesse grupo, observa-se, ainda, que os sujeitos foram mais lentos para atribuir gênero a pseudopalavras com terminações em *-e*, *-ume* e *-ade*, (1025ms e 1016ms) que a itens com terminação em *-agem* (989ms), indicando um efeito inibidor da vogal final átona *-e*, que pode ser associada tanto ao gênero feminino quanto ao gênero masculino. Entre os grupo das irregulares, surpreendentemente, os sujeitos foram quase tão rápidos, em média, para atribuir gênero a pseudopalavras terminadas em *-ema* quanto a pseudopalavras terminadas em *-a* (934ms vs. 953ms), sugerindo que, embora *-ema* seja tipicamente masculina, a vogal *-a* átona final exerce maior poder associativo que a terminação como um todo. Em contrapartida, observa-se que os sujeitos foram mais lentos para atribuir gênero aos itens terminados em *-ão* que a itens terminados em *-o* (1022 vs. 908), revelando o efeito inibidor dessa terminação.

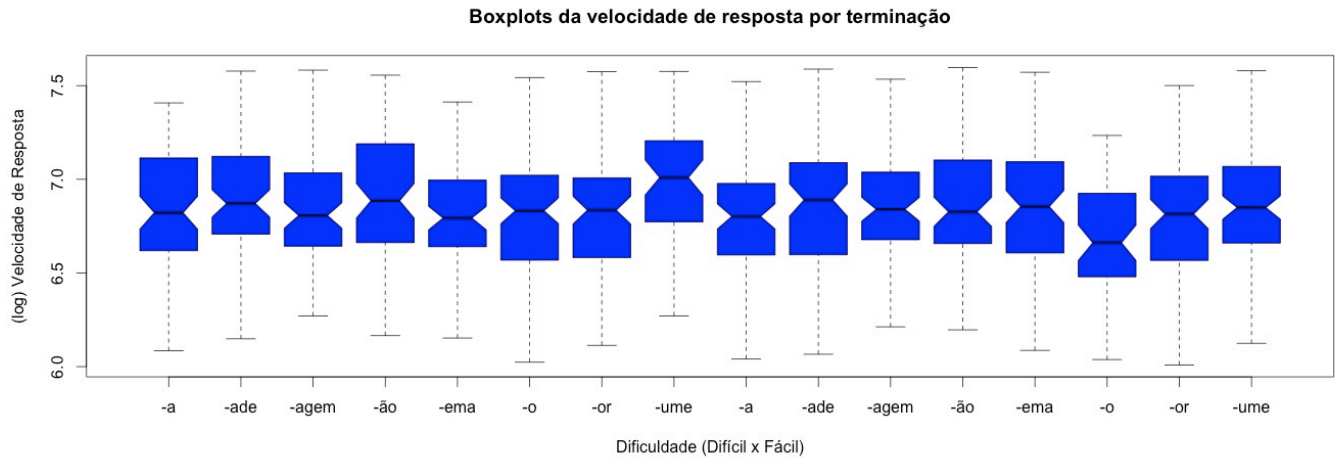
Em relação ao processamento da atribuição de gênero a pseudopalavras difíceis *versus* pseudopalavras fáceis, a inspeção visual da Tabela 9, abaixo, revela um efeito facilitador da estrutura fônica das pseudopalavras fáceis em todas as categorias, com exceção das pseudopalavras terminadas em *-ema*, do grupo da irregulares, e terminação *-agem*, do grupo das Transparentes.

Tabela 9: Descrição da velocidade de resposta de 28 sujeitos por terminação: fáceis vs. difíceis. DP: desvio padrão.

Terminação da pseudopalavra	Fáceis			Difíceis		
	Média	Mediana	DP	Média	Mediana	DP
-a	905ms	898ms	288 ms	964ms	913ms	319 ms
-o	834ms	780ms	247 ms	945ms	921ms	312 ms
-agem	989ms	932ms	314 ms	968ms	902ms	295 ms
-ade	1003ms	978ms	351 ms	1047ms	962ms	349 ms
-ume	976ms	910ms	294 ms	1096ms	1071ms	357 ms
-or	930ms	910ms	280 ms	949ms	913ms	311 ms
-ema	979ms	946ms	312 ms	927ms	867ms	276 ms
-ão	1006ms	909ms	328 ms	1038ms	977ms	336 ms

Os *Diagramas de Caixa* da Figura 6, a seguir, ilustram o efeito do grau de dificuldade da pseudopalavras.

Figura 6: *Diagramas de Caixa* da velocidade de resposta de 28 sujeitos em uma tarefa de atribuição de gênero a 48 pseudopalavras distribuídas em 2 grupos:



grupo das difíceis à esquerda e grupo das fáceis à direita.

A Tabela 10, abaixo, apresenta as variáveis incluídas no modelo que melhor explicam a velocidade de resposta de 28 sujeitos no Experimento 3. Assim como no Experimento 1 e Experimento 2, as análises estatísticas dos dados obtidos no Experimento 3 foram executadas em linguagem *R* e baseadas na transformação logarítmica (log natural) da variável dependente (velocidade de resposta). Valores com desvio padrão de 2,5 pontos abaixo e acima da média (*outliers*) foram removidos do conjunto de dados (2,2% dos dados).

Para implementação do MEM, usei o pacote *lme4* (BATES *et al.*, 2014). A Tabela 10, abaixo, apresenta os resultados obtidos pelo MEM na análise do processamento da atribuição de gênero às pseudopalavras do Experimento 3.

Tabela 10: Variáveis incluídas no modelo que melhor explicam a velocidade de resposta de 28 sujeitos no Experimento 3.

Efeitos fixos	Estimativa	Erro padrão	t-valor	p-valor
<i>intercept</i>	6.796437	0.085841	79.17	p< 0,01
Forma	0.073507	0.036954	1.99	p<0,05
Terminação	-0.093852	0.036970	-2.54	p<0,05

Dificuldade	-0.037386	0.018631	-2.01	p<0,05
Efeitos Aleatórios	Nome	Variância	Desvio Padrão	
Sujeitos	<i>intercept</i>	0.039755	0.19939	
Estímulos	<i>intercept</i>	0.001905	0.04364	
	Resíduo	0.058706	0.24229	

Após a limpeza do conjunto de dados, seis diferentes MEM foram construídos, usando o pacote *lme4* (BATES *et al.*, 2014). O primeiro modelo não incluía interação entre as variáveis fixas (*fixed effects*) e, nos outros cinco modelos, interações entre as variáveis *Forma*, *Dificuldade* e *Forma*, *Dificuldade* e *Sexo* foram incluídas. A significância das interações incluídas foi verificada por meio do *likelihood ratio test*, pacote *stats* (R DEVELOPMENT CORE TEAM, 2013). Os valores de *p* foram obtidos através do teste ANOVA, tipo 3, teste 'Chisq' (qui quadrado). A inspeção visual do gráfico de distribuição de resíduos do modelo escolhido não revelou anormalidade na distribuição dos resíduos.

Na comparação dos modelos, o *likelihood ratio test* mostrou que nenhuma das interações incluídas no modelo era estatisticamente significativa ($p < 0.05$). O modelo que melhor descreve a velocidade de resposta dos sujeitos na atribuição de gênero revelou um efeito do fator *Forma* ($t = 1,99$, $p = 0,02$) atuando sobre o processamento, bem como um efeito da *Terminação* ($t = -2,54$, $p = 0,03$). O fator *Dificuldade* mostrou-se também significativo ($t = -2,01$, $p = 0,04$). O efeito individual do fator *Dificuldade* e a ausência de interação entre os fatores *Terminação* e *Dificuldade* indicam que este o fator *Dificuldade* influenciou o processamento de todas as pseudopalavras independentemente de sua terminação. Figura 7 e Figura 8, a seguir, ilustram o efeito do fator *Dificuldade*.

Figura 7: Gráfico de interação entre o fator Dificuldade (fáceis vs. difíceis) e velocidade de resposta por terminação da pseudopalavra.

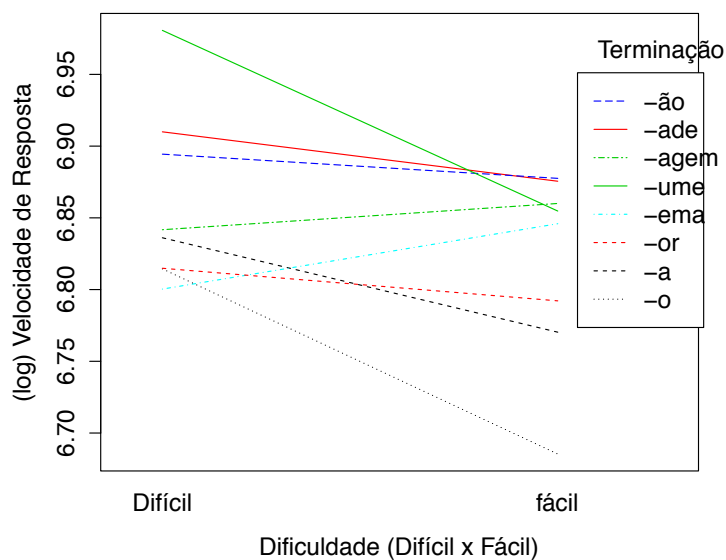
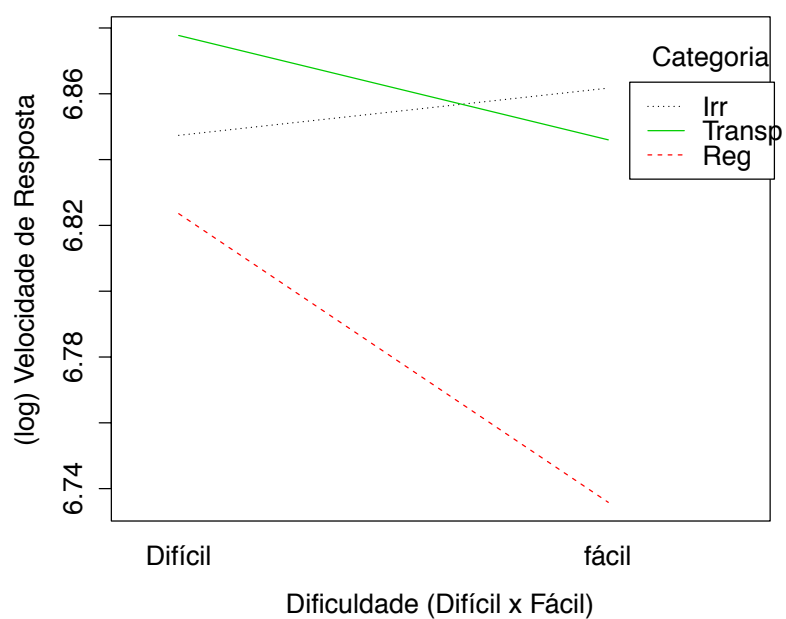


Figura 8: Gráfico de interação entre o fator Dificuldade (fáceis vs. difíceis) e velocidade de resposta por categoria de pseudopalavra.



4.4.3 Discussão dos resultados

Nesta seção, discuto primeiramente os resultados obtidos na modelagem dos dados (conforme MEM) e, a seguir, discuto os efeitos observados para cada uma das terminações incluídas no Experimento 3. Esses efeitos estão ilustrados nas Figuras 7 e Figura 8 acima.

Como esperado, o fator *Dificuldade* aparece como relevante, ao passo que a interação desse fator com o grupo das pseudopalavras de terminações irregulares ou transparentes não foi estatisticamente significativa ($p > 0,05$), sugerindo que o processamento tanto de itens com terminações regulares (-a e -o) quanto terminações irregulares (-ema, -ão) ou transparentes (-agem, -ade, -ume, -or) foi afetado pelo grau de dificuldade de associação da pseudopalavra.

O efeito individual dos fatores *Terminação* e *Forma* indicam que o processamento de todas as categorias de forma de pseudopalavras e terminações são diferentes entre si. Novamente, nesse experimento, o fator *Sexo* não foi estatisticamente significativo, bem como não alcançou significância a interação entre os fatores *Sexo* e *Forma*. Esse resultado indica que diferenças de sexo não impactaram o processamento do gênero, conforme assumem os modelos de via dual (cf. Cap.3). Da mesma forma, as variáveis controle incluídas no MEM, *Ordem* de apresentação da tarefas e número de *Grafemas* também não foram significativas, portanto, assim como no Experimento 2, o processamento do gênero no Experimento 3 não sofreu interferência de fatores relacionados ao desenho experimental.

Com relação aos efeitos exibidos nas Figuras 7 e 8 e os dados apresentados na Tabela 7 de classificação das pseudopalavras em classes de gênero, é possível notar que, como esperado, o efeito do fator *Dificuldade* foi observado na velocidade de resposta dos sujeitos para atribuição de gênero aos itens de quase todas as terminações. O resultado inesperado foi o efeito inverso observado para -ema e -agem (Figura 8). No entanto, em termos gerais, os sujeitos foram mais rápidos para atribuir gênero aos itens de fácil associação que aos de difícil associação. Esse efeito se fez presente, também, na classificação dos itens em categorias de gênero (cf. Tabela 7), confirmando a hipótese inicial de que pseudopalavras com terminação típicas do gênero feminino seriam classificadas, em sua maioria, como femininas (exceto para a terminação -agem, discutida a seguir); pseudopalavras com

terminações típicas do gênero masculino seriam classificadas, em sua maioria, como masculinas. Conforme previsto (cf. Seção 4.4), essas porcentagens aumentaram quando as pseudopalavras eram de fácil associação, sugerindo, novamente o efeito da força associativa dos itens fáceis.

Detenho-me, agora, na discussão do efeito inverso (cf. Figura 8) observado no processamento da atribuição de gênero aos itens terminados em *-ema* e, a seguir, *-agem*. Como se pode observar na Figura 8, os sujeitos foram mais rápidos para atribuir gênero a pseudopalavras difíceis terminadas em *-ema* (*cetema*) do que a pseudopalavras fáceis com essa terminação (*Sisfema*). Esse efeito inverso pode ser explicado se olharmos para os dados de classificação dos itens em categorias de gênero (Tabela 7). As porcentagens de classificação mostram que, os itens difíceis terminados em *-ema* foram, em sua maioria, classificados como femininos (71,4%). Essa porcentagem diminui quando os itens são fáceis (61,9%). Esse resultado pode ser explicado com base no fato de os itens fáceis compartilharem traços estruturais com palavras do português do gênero masculino. Esse resultado sugere que os sujeitos processaram grande parte dos itens difíceis de forma a decompô-los para fins de aplicação da regra da terminação (já que a terminação em *-a* para atribuição de gênero feminino é produtiva), ao passo que, no processamento da atribuição de gênero a itens fáceis, uma força associativa parece atuar. Uma evidência da atuação dessa força é o aumento na velocidade de resposta dos sujeitos para itens fáceis bem como o declínio das porcentagens de classificação desses itens como femininos. Como observado no Experimento 2, o custo do processamento para itens irregulares quanto ao gênero (femininos terminados em *-o*, e masculinos terminados em *-a*) é maior em comparação a todas as outras terminações (*-agem*, *-ade*, *-or*, *-ume*, *-a*, *-o*). Portanto, a hipótese que levanto é: ao associar as pseudopalavras fáceis terminadas em *-ema* com a estrutura fônica de seu par mínimo hipotético, um processo associativo foi ativado. Esse processo sofreu o impacto de forças em competição: de um lado a força da terminação *-a*, induzindo a atribuição do gênero feminino e, de outro, a força associativa da pseudopalavra que se assemelha a uma palavra real do português do gênero masculino. Isso explicaria um processamento mais rápido dos itens difíceis em oposição a um processamento mais lento da atribuição de gênero aos itens fáceis

Cabe ressaltar, ainda, que o efeito inverso acima discutido sugere que os sujeitos decompueram os itens difíceis e ativaram um mecanismo associativo no processamento da atribuição de gênero aos itens fáceis. Suponho que esses dois processos tenham atuado de forma paralela (decomposição das difíceis e associação das fáceis) também em relação ao processamento da atribuição de gênero às pseudopalavras de outras terminações.

Uma questão que merece destaque é: Por que esse efeito inverso não foi observado no processamento dos itens terminados em *-ão*, já que, como se verifica em *-ema*, essa terminação pode ser associada tanto ao gênero masculino quanto do gênero feminino? uma possível resposta para essa pergunta diz respeito às características distribucionais das terminações *-ema* e *-ão*. Enquanto a terminação *-ema* é infrequente no português, *-ão* é altamente frequente. No *Corpus Brasileiro*, que contém aproximadamente mais um bilhão de palavras, substantivos terminados em *-ema* aparecem apenas 1254 vezes. Esse número aumenta consideravelmente quando a terminação é *-ão*, que apareceu 51114 vezes, sendo que a distribuição de itens com essa terminação entre as categorias de gênero (feminino e masculino) favorece a categoria feminina. Portanto, pode-se concluir que: Em um embate de forças, a força associativa de *-ema* para o gênero masculino é menor que a força da associação da vogal átona final *-a* dessa mesma terminação para o gênero feminino. O mesmo raciocínio se aplica à terminação em *-ão*: num embate de forças, a vogal átona final *-o* para o gênero masculino vence sobre a terminação *-ão* para o gênero feminino. Esse parece ser o caso do processamento dos itens difíceis, seja este com terminação em *-ema* ou terminação em *-ão*. No entanto, quando os sujeitos foram expostos à pseudopalavras de fácil associação com essas terminações, um processo de desambiguação eclodiu. A questão é que o processo de desambiguação do gênero foi mais fácil para os pares mínimos hipotéticos terminados *-ão* pelo fato de essa terminação ser mais frequente e, além disso, mais frequentemente associada ao gênero feminino, fato que não se aplica à terminação *-ema*. Apesar da terminação *-ema* ser tipicamente masculina, como já mencionado, sua frequência é muita baixa. Portanto, no processamento da atribuição de gênero aos itens terminados em *-ema* verifica-se um processo competitivo mais longo, o que acabou induzindo menos eficiência (menos velocidade) dos sujeitos para atribuir gênero a esses itens, provocando, assim, o efeito inverso observado.

No que se refere ao processamento da atribuição de gênero a pseudopalavras terminadas em *-agem*, a inspeção visual da Figura 7 revela que o efeito inverso é pequeno para esses itens. Além disso, as porcentagens de classificação dessas pseudopalavras no gênero feminino é idêntico entre itens fáceis (84,5%) e difíceis (84,5%), sugerindo a decomposição de todas as formas. Suponho que a decomposição tenha ocorrido devido a semelhança estrutural dos itens terminados em *-agem* incluídos no experimento (cf. Apêndice D).

Em resumo, como já mencionado, os itens de fácil associação do presente experimento são pares mínimos hipotéticos de palavras reais do português (ex: *Família / Bamília*). Assim sendo, essas pseudopalavras apresentam uma força associativa maior que as pseudopalavras difíceis em virtude de sua estrutura fônica estar em conexão com estruturas fonológicas pré-existentes no léxico mental do falantes do português. De acordo com o *Network Model*, é possível observar uma redução no custo do processamento de itens que estão em conexão. O decréscimo na velocidade de resposta dos sujeitos para o processamento da atribuição de gênero para esse grupo reflete essa redução. No caso das pseudopalavras terminadas em *-ema*, essa redução não foi observada em virtude da ambiguidade dessa terminação bem como sua infrequência de sua associação com o gênero masculino. Em acréscimo, os resultados obtidos para o processamento da atribuição de gênero às pseudopalavras terminadas em *-ema* sugerem, também, que o processamento decomposicional para atribuição de gênero ficou restrito ao grupo de pseudopalavras de difícil decomposição, exceto no caso da terminação *-agem* em que a decomposição parece atuar em todos os itens em virtude de suas semelhanças estruturais. No entanto, esse resultado não põe em xeque as hipótese levantadas para o presente experimento. Em termos gerais, os resultados aqui obtidos sugerem que, mais que a simples decomposição das formas para a aplicação da regra da terminação, o que parece estar atuando no processamento é a força associativa das propriedades estruturais dos itens como um todo. Se, conforme afirmam os modelos de via dual, o processamento das regularidades fosse de natureza computacional, ao confrontar o processamento dos itens fáceis com o processamento dos itens difíceis, o efeito do fator *Dificuldade* não teria sido observado.

4.5 Discussão geral do capítulo

Neste capítulo, apresentei três experimentos comportamentais que tiveram como objetivo principal investigar se o processamento do gênero gramatical é governado por um sistema dual (em que se observa, de um lado, aplicação de regras para formas regulares e, de outro, acesso de itens irregulares e subregulares na memória) ou governado por um único sistema de memória de natureza associativa.

Há evidências nos dados que dão suporte à hipótese de um processamento dual do gênero gramatical do português. No Experimento 1, o efeito do fator *Forma* indica diferenças entre os processamentos de itens regulares e opacos. O efeito da interação entre os fatores *Forma* e *Frequência* indica que a frequência de ocorrência dos itens influenciou as formas regulares e opacas de maneira distinta. Por outro lado, em uma análise exploratória dos dados, o teste *t* pareado mostrou um efeito da frequência atuando em ambas as categorias de forma, sendo o efeito da frequência maior para o grupo das regulares. Esse resultado indica, portanto, que independentemente de sua forma, todos os itens foram extraídos da memória. Indicam, ainda, que as regulares infrequentes ativaram um mecanismo computacional, já que os participantes foram mais lentos para executar a concordância de gênero com esses itens.

Há também evidências nos dados que suportam a hipótese do processamento do gênero gramatical do português ser governado por um único sistema associativo. O efeito independente da frequência presente no Experimento 2 revela que todos os itens, sejam eles Regulares, Irregulares, Transparentes ou Opacos quanto ao gênero, foram acessados na memória para o processamento do seu gênero. Além disso, a ausência de interação entre os fatores *Frequência* e *Forma* sugere que o tamanho do efeito da frequência foi similar entre todas as categorias de palavras. O efeito da *Forma* nesse experimento emergiu como significativo tanto para a categoria das formas regulares quanto para a categoria das formas transparentes. Formas regulares, entretanto, emergiram como estatisticamente mais relevantes que transparentes ($p < 0,01$ vs. $p \leq 0,05$). Esses resultados sugerem que i) tanto a regularidade quanto a transparência atuam como facilitadores no processamento da atribuição de gênero, já que a velocidade dos sujeitos para ambas as categorias foi inferior à velocidade na atribuição do gênero às categorias de itens opacos e irregulares; ii) em oposição às premissas dos modelos de via dual que propõem um sistema de processamento das regularidades

distinto das demais formas, no Experimento 2, os resultados mostram que o processamento de formas transparentes se aproxima do processamento de formas regulares e, por fim, que iii) a força da representação das terminações *-a* e *-o* é superior à força da representação de outras terminações típicas do feminino e masculino para a classificação de palavras em categorias de gênero.

A força associativa das terminações *-a* e *-o* ficou ainda mais evidente no Experimento 3. Nesse último experimento, foi possível verificar que, quando uma terminação podia estar associada tanto ao gênero feminino quanto ao gênero masculino (*-ema* e *-ão*), os sujeitos eram mais sensíveis à vogal átona final *-a* e final *-o* que às terminações *-ema* e *-ão* como um todo. No entanto, a desambiguação do gênero dessas terminações se fez presente quando a pseudopalavra compartilhava traços estruturais (um par mínimo hipotético) com uma forma pré-existente no léxico mental dos falantes. Esse resultado evidencia, portanto, a atuação de um processo associativo.

No Experimento 3, o MEM revelou também um efeito do fator *Terminação*, indicando a terminação exerce um impacto no processamento do gênero. Por outro lado, o efeito independente e estatisticamente relevante do fator *Dificuldade* nesse experimento deixou claro que a atuação do processo associativo para o processamento do gênero não se limita às terminações ambíguas (categoria das irregulares), mas exerce um impacto, também, no processamento do gênero de pseudopalavras de terminação regular e transparente. Esses dados indicam que a memória associativa exerce papel significativo no processamento do gênero. Indicam, ainda, que os substantivos precisam ser armazenados para que se tornem itens exemplares de um processo associativo (cf. Bybee, 1995).

Apesar de estudos como o de Ullman *et al.* (2002) apontar para diferenças no processamento entre homens e mulheres, os dados aqui apresentados não fornecem evidências para diferenças entre os sexos, já que o fator *Sexo* não interagiu com o fator *Forma* e *Frequência* do itens em nenhum dos três experimentos. Além disso, evidências para a hipótese de que formas regulares são computadas *online* também não foram encontradas. Como já mencionado, no Experimento 1, efeitos de frequência foram encontrados para a concordância de gênero com formas regulares (substantivos femininos terminadas em *-a*, e masculinos em *-o*) bem como para palavras opacas quanto ao gênero. No Experimento 2, esse efeito da frequência atingiu todas as categorias

de formas e, no Experimento 3, foi possível observar uma diferença no processamento de pseudopalavras difíceis e fáceis com terminação regular quanto ao gênero do português.

Em resumo, ao processar o gênero gramatical os falantes ativam um sistema computacional para terminações regulares? Levando em conta os dados obtidos nos três experimentos apresentados neste capítulo, considero que um processamento em duas rotas é improvável. Os resultados apontam para uma atuação de uma memória de natureza associativa em todas as categorias de forma. Se assim for, um modelo conexionista será capaz de reproduzir os resultados dos estudos aqui apresentados. No capítulo 5, a seguir, descrevo a construção de um modelo computacional de base conexionista que simula o processamento do gênero português. Esse modelo simula a atribuição de gênero a substantivos com terminação ambígua, em especial, a atribuição de gênero a substantivos terminados em *-ema* e *-e*. Essa escolha está pautada na observação de que, a desambiguação do gênero de substantivos com essa terminação, ao invés de regra, o modelo precisa levar em consideração a estrutura das palavras como um todo (os traços compartilhados entre as estruturas) e não apenas sua terminação, assim como ocorre em um processo associativo.

CAPÍTULO 5

5 CLASSIFICAÇÃO DE SUBSTANTIVOS DO PORTUGUÊS EM CATEGORIAS DE GÊNERO: UMA ABORDAGEM CONEXIONISTA

5.1 Introdução

A inteligência artificial é um campo da ciência que lida com a possibilidade de criar seres artificiais, isto é, construir máquinas que possam pensar, aprender e criar assim como seres humanos (RABUÑAL; DORADO, 2006). Essa área da ciência está dividida em dois paradigmas teóricos principais: o paradigma conexionista e o paradigma simbólico.

Diferentemente do paradigma simbólico que tem como objetivo representar o conhecimento por meio de codificação de regras usando uma linguagem formal, a inteligência artificial conexionista parte da ideia de que os processos mentais podem ser modelados por uma rede de unidades neurais interconectadas. Nesse sentido, a arquitetura de um modelo computacional conexionista é inspirada na arquitetura do sistema neural humano e, por isso, compreende neurônios artificiais organizados em uma rede de conexões. Esses modelos estão em consonância com a regra de Hebb a qual postula que a conexão entre os neurônios é reforçada todas as vezes que é usada (MCCULLOCH; PITTS, 1943).

Sistemas construídos com base no paradigma conexionista são capazes de aprender por analogia padrões implícitos presentes em um conjunto de dados e reconhecê-los em bases de dados diferentes, isto é, em um conjunto de dados que não fizeram parte do processo de aprendizagem do modelo, sem que regras sejam explicitamente ensinadas (cf. PLUNKETT; MARCHMAN, 1991).

Ao longo das últimas décadas, com o avanço da ciência e tecnologia no desenvolvimento de computadores com maior capacidade de processamento e armazenamento de informação, os modelos conexionistas têm se tornado cada vez mais sofisticados e úteis em diferentes campos do conhecimento. A área de aquisição de linguagem, por exemplo, tem se beneficiado dos recursos que esses sistemas inteligentes podem fornecer para a simulação da aquisição e do processamento mental de certos aspectos da linguagem humana.

Vários modelos conexionistas concentraram-se na simulação do processamento e aquisição da morfologia flexional das línguas naturais. O modelo conexionista implementado por Rumelhart e McClellands (1986), por exemplo, simulou as três fases do processo de aquisição por

crianças da morfologia flexional do passado do inglês como língua materna. De acordo com autores, na fase 1, crianças adquirem um número reduzido de verbos regulares e irregulares de alta frequência. Porém, nessa fase inicial de aquisição, regras de produção do passado ainda não foram adquiridas. Na fase 2, por sua vez, regras morfológicas começam a emergir. É nesse fase que erros de produção, como regularização de verbos irregulares (por exemplo: *come/*comed*) aparecem. Já na fase 3, verbos regulares e irregulares coexistem no léxico mental das crianças. Verbos irregulares são produzidos corretamente e regras são aplicadas na produção de verbos regulares. Na simulação da fase 1, o modelo reproduziu os estágios iniciais de aquisição da morfologia flexional do passado do inglês. Assim como observado em crianças, o modelo connexionista não revelou diferenças na simulação da flexão de verbos regulares e irregulares. A simulação da fase 2 revelou que o modelo regularizava verbos irregulares, reproduzindo novamente um padrão observado na fase 2 do processo de aquisição da morfologia flexional verbal do passado. A simulação da fase 3, por sua vez, revelou que o modelo é capaz de reconhecer os padrões estruturais de verbos regulares e irregulares nunca apresentados anteriormente à rede. Esses resultados sugerem que um modelo connexionista é eficiente em reproduzir o comportamento de crianças durante as fases de aquisição da morfologia verbal do passado do inglês. Sugerem, ainda, a sensibilidade do modelos connexionistas em relação aos padrões de regularidade bem como subregularidades da morfologia verbal do passado do inglês. Esses resultados evidenciam que tanto itens regulares quanto itens irregulares são passíveis de serem processados por um sistema de via única sem que regras sejam explicitamente ensinadas.

Além da morfologia verbal, outros modelos abordaram questões relativas à aquisição e processamento da morfologia de número. Plunkett e Juola (1999) modelaram a aquisição da morfologia verbal do passado do inglês, bem como a aquisição da morfologia de número regular e irregular da língua. Novamente, os resultados dessas simulações demonstraram que modelos connexionistas são capazes de reproduzir o comportamento de crianças durante o processo de aquisição desses aspectos da morfologia flexional.

No que concerne aos modelos que simularam o processamento do gênero gramatical é possível fazer alusão ao trabalho de Mirkovic, MacDonald e Seidenberg (2004) que compararam o desempenho de três modelos os quais se concentraram em questões relativas ao processamento do gênero. O propósito do estudo foi investigar se o gênero em língua sérvia está associado à informação marcada na

estrutura fonológica dos substantivos e a informações semânticas. Em outras palavras, Mirkovic, MacDonald e Seidenberg (2004) investigaram, por meio de modelos conexionistas, se o gênero em sérvio emerge de regularidades semânticas (nomes animados vs. inanimados) e fonológicas (nomes marcados ou não quanto ao gênero). Os resultados mostraram que existe uma correlação alta de fatores semânticos (animacidade) e as categorias de gênero. Além disso, o melhor desempenho foi atingido pelo modelo que incluía informações explícitas de marcação de gênero na estrutura fonológica das palavras. Com base nesses resultados, os pesquisadores concluíram que os modelos conexionistas empregados fornecem dados preliminares consistentes com a hipótese de que o gênero gramatical pode ser aprendido e representado por um único sistema em termos de regularidades morfofonológicas e semânticas (MIRKOVIC, MACDONALD E SEIDENBERG, 2004, p.23).

Em um trabalho posterior, Mirkovic, Seidenberg e Joannis (2011) investigaram, também por meio de uma rede neural artificial, o papel que fatores semânticos e fonológicos cumprem na representação mental do gênero em sérvio. O segundo objetivo desse trabalho foi usar a rede neural como uma ferramenta para a investigação de padrões no sistema flexional do sérvio de difícil identificação. Como *input*, o arquivo de treino recebeu informações de lema, número, caso e gênero para que o modelo fosse capaz de produzir a forma fonológica correta dos nomes. O modelo foi capaz de extrair padrões de uma grande quantidade de formas flexionadas (por volta de 3000 nomes substantivos) e fornecer informações sobre a complexidade do sistema do sérvio.

É importante salientar que esses modelos lidam com o aprendizado de máquina de tipo supervisionado (do inglês: *supervised machine learning*). Um aprendizado de máquina supervisionado tem como objetivo aprender padrões para ser capaz de identificá-los em bases de dados diversas. Esse tipo de aprendizado compreende duas fases: a fase de treino e a fase de teste (simulação da rede). Na fase de treino, a rede neural artificial ‘aprende’ por analogia o padrão existente na relação entre *inputs* e *outputs* de uma base de dados em ciclos de propagação/adaptação. A fase de teste consiste na apresentação de um conjunto de dados não visto pelo algoritmo na fase de treino para que se possa avaliar o seu desempenho em relação às generalizações que ele é capaz de realizar após o treinamento. Em linhas gerais, nessa fase, o algoritmo compara a sua resposta (seus *outputs*) com os *outputs* desejados. Grande parte dos estudos que tiveram como objetivo modelar

fases da aquisição e do processamento de aspectos da linguagem usaram uma rede neural multicamadas (*perceptron* multicamadas) de tipo *feedforward* e treinamento usando o algoritmo *backpropagation* (PLUNKETT, MARCHMAN, 1991, 1993; PLUNKETT, JUOLA, 1999; MACWHINNEY; LEINBACH, 1991).

Uma rede *feedforward* é chamada de multicamadas pelo fato de apresentar uma camada de entrada representando os *inputs* da rede, uma (ou mais) camadas intermediárias (*hidden layers*) compostas por um ou mais neurônios (*hidden units*) e uma camada de saída (*output*) que também por ser composta por um ou mais neurônios. Em uma rede *feedforward*, os sinais de *input* seguem em direção ao *output*. Com o treinamento usando o algoritmo *backpropagation*, os erros dos elementos processadores da camada de saída são retro-propagados para as camadas intermediárias. Cada elemento (cada neurônio artificial) da camada intermediária recebe uma porção do erro total. Com base nos sinais de erro recebidos, os pesos das conexões são ajustados para cada elemento de modo a fazer a rede convergir (encontrar uma solução) para um modelo final que dá conta de todos os padrões presentes no conjunto de dados de treinamento.

Os modelos conexionistas com foco na simulação da aquisição e processamento de aspectos da linguagem humana demonstraram eficiência no aprendizado de padrões estruturais de formas linguísticas. Esses resultados mostraram que as redes neurais são capazes de generalizar o conhecimento adquirido para novos exemplos. Tendo em vista a eficiência dos modelos conexionistas para a simulação da aquisição e do processamento de aspectos da linguagem humana, no presente capítulo, o processamento do gênero gramatical do português será modelado sob a perspectiva conexionista. Assim como os estudos que modelaram as fases da aquisição e do processamento de aspectos da morfologia flexional das línguas naturais, o experimento aqui apresentado adota um aprendizado de máquina supervisionado para a classificação de nomes substantivos do português em categorias de gênero: masculino ou feminino.

O objetivo geral dessa modelagem foi investigar se os resultados observados nos estudos psicolinguísticos apresentados no Capítulo 4 desta tese podem ser explicados por um modelo conexionista. Como visto na Introdução e no Capítulo 3 dessa tese, a visão conexionista de processamento da linguagem, em oposição à visão dual, rejeita a hipótese da dissociação cognitiva entre formas regulares e irregulares. Modelos conexionistas assumem que todas as formas são processadas por um único sistema de memória associativa. Assim sendo, alguns dos

padrões de processamento observados nos resultados experimentos comportamentais reportados no Capítulo 4 forem reproduzidos pelo modelo conexcionista, então é possível assumir que um sistema de via única é capaz de explicar aspectos do processamento do gênero gramatical do português. Mais especificamente, busco responder às seguintes questões:

- 1) Os resultados dos Experimentos 1 e Experimento 2 do Capítulo 4 desta tese mostraram que os sujeitos foram sensíveis à frequência de ocorrência dos padrões estruturais dos itens lexicais apresentados para classificação em categorias de gênero. Em que medida uma rede neural artificial é também sensível à frequência dos itens e suas propriedades estruturais?
- 2) Os resultados do Experimento 3 mostraram que o processo associativo para atribuição de gênero não se restringe apenas à terminação dos vocábulos, mas levam em conta os traços compartilhados com outras formas presentes no léxico mental dos falantes. A generalização realizada por uma rede neural para classificação de formas em categorias de gênero será baseada apenas na relação existente entre a terminação do vocábulo e seu gênero ou a estrutura como um todo das palavras presentes no treinamento serão consideradas pela rede neural em seu *output*?

Como já mencionado, para avaliar o desempenho da rede é preciso primeiramente treiná-la e, então, testá-la. No presente estudo, para a fase de teste, tarefas de atribuição de gênero à palavras e pseudopalavras serão simuladas. Assim como as tarefas de atribuição de gênero empregadas Capítulo 4 desta tese que contou com a participação de falantes nativos do português, a rede deve atribuir o gênero feminino ou masculino a palavras e pseudopalavras a ela apresentadas com base no conhecimento adquirido na fase de treino.

Nas próximas seções, os critérios adotados na construção do conjunto de dados para o treinamento da rede, a descrição de cada uma das simulações e os resultados de cada uma delas serão apresentados. Por fim, discuto os resultados.

5.2 A construção do corpus

O *corpus* para as fases de treinamento e teste da rede neural foi

organizado em uma matriz em formato csv (*comma separated values*) contendo nove colunas e 2386 linhas (instâncias). Cada coluna da matriz correspondia a posições dos grafemas de um substantivo: grafema 1, grafema 2, grafema 3, grafema 4, grafema 5, grafema 6, grafema 7, grafema 8 e grafema 9. Cada linha da matriz correspondia aos valores que preenchem a posição dos grafemas de um substantivo (letras do alfabeto). Portanto, em cada linha (instância) um substantivo foi introduzido. Cada um dos grafemas foi representado por um número. A Tabela 11, abaixo, apresenta os valores de cada grafema dos substantivos presentes no corpus.

Tabela 11: Representação numérica dos grafemas dos substantivos presentes no corpus.

Grafemas	Valores
<i>a</i>	1
<i>ã</i>	21
<i>á</i>	25
<i>e</i>	4
<i>i</i>	8
<i>í</i>	22
<i>o</i>	12
<i>ó</i>	23
<i>u</i>	18
<i>ú</i>	29
<i>c</i>	2
<i>d</i>	3
<i>f</i>	5
<i>g</i>	6
<i>h</i>	7
<i>l</i>	9
<i>m</i>	10
<i>n</i>	11
<i>p</i>	13
<i>q</i>	14
<i>r</i>	15
<i>s</i>	16
<i>t</i>	17
<i>v</i>	19
<i>z</i>	20

As posições dos grafemas de cada substantivo e seus valores representaram os *inputs* apresentados à rede. Esse procedimento

permitiu a simulação da leitura de cada substantivo na medida em que a rede "aprendeu" que cada um deles é representado por uma sequência de grafemas (máximo nove grafemas) com valores distintos.

Caso um determinado substantivo apresentasse uma quantidade de grafemas inferior a nove, a coluna correspondente ao valor do grafema foi preenchida com valor "0". A anotação de cada substantivo foi realizada de trás para frente (do último grafema ao primeiro) para que fosse possível que o grafemas finais de todos os substantivos estivessem alinhados em uma mesma coluna. Esse procedimento foi importante para que a rede "compreendesse" que o último *input* correspondia à terminação dos substantivos. Os *inputs* das 2386 instâncias foram apresentados à rede de uma só vez (aleatoriamente) para simular a apresentação visual de estímulos.

Os substantivos introduzidos no *corpus* foram selecionados do conjunto de estímulos criados para o Experimento 2 do Capítulo 4 (cf. Apêndice E). No total, foram selecionados 40 diferentes substantivos do português distribuídos em 4 categorias de forma. A Tabela 12, abaixo, apresenta a distribuição.

Tabela 12: Categorias de formas e quantidade de substantivos escolhidos para cada categoria.

Categorias de forma	Terminação	Quantidade de substantivos	Exemplos
Regular	<i>-a</i>	11	<i>Linha, ideia</i>
	<i>-o</i>	11	<i>Carro, acordo</i>
Irregular	<i>-ão</i>	4	<i>Mão, região</i>
	<i>-ema</i>	3	<i>Cinema, sistema</i>
Transparente	<i>-agem</i>	2	<i>Viagem, Imagem</i>
	<i>-or</i>	4	<i>Setor, valor</i>

Opacas	-e	5 (3 masculinas e 2 femininas)	<i>Filme, nome, saude</i>
--------	----	--------------------------------------	-------------------------------

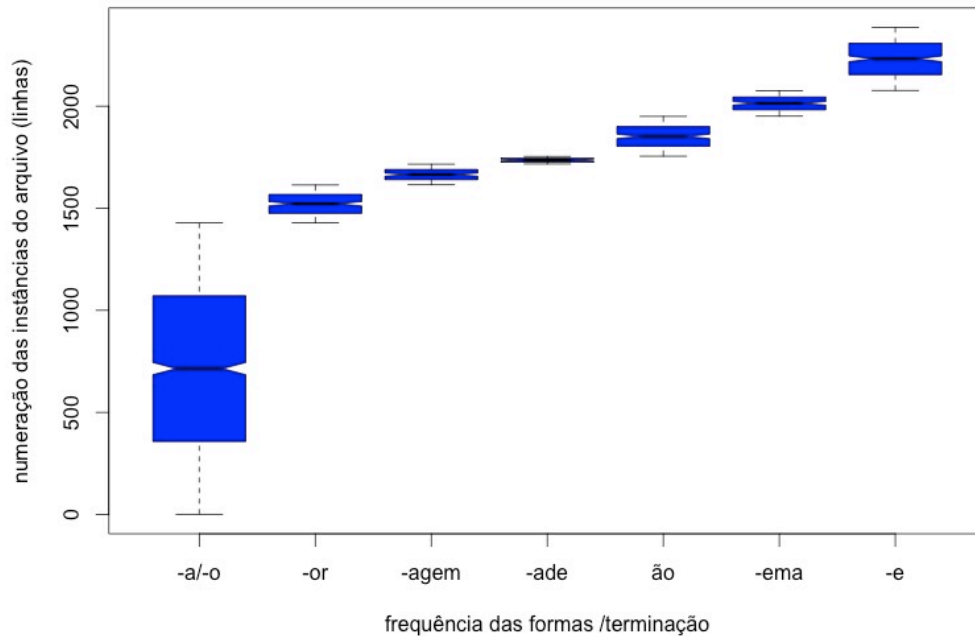
Note que a categoria dos substantivos regulares representou mais de 50% (22 de 40) das formas selecionadas para a composição do *corpus*. Esse procedimento foi adotado para simular a proporção maior de itens regulares quanto ao gênero no português em relação às outras formas (cf. Cap.2). Cabe esclarecer, também, que o *corpus* foi construído de maneira a compreender substantivos que apresentam e que não apresentam uma relação entre estrutura e gênero. Por exemplo, foram apresentados três substantivos masculinos com terminação em *-e*, e dois femininos também terminadas em *-e*. Além disso, o *corpus* foi composto por substantivos masculinos e femininos com terminação em *-a*, masculinos e femininos com terminação em *-ão*.

Cada um dos substantivos entrou no arquivo de treino uma quantidade de vezes proporcional (10%) à sua frequência de ocorrência no *Corpus Brasileiro*. Por exemplo, se a frequência de um substantivo era de 900 por milhão no *Corpus Brasileiro*, então esse substantivo foi apresentado à rede 90 vezes. Foram selecionados substantivos de frequência média e alta (cf. Apêndice E). O arquivo final compreendeu 2386 entradas de substantivos (2386 instâncias).

O *output* do modelo foi representado por valores categóricos em dois níveis: ‘F’ para o gênero feminino, e ‘M’ para o gênero masculino.

Os *diagramas de caixa* da Figura 9 ilustram a quantidade de entrada de formas e a organização dessas categorias de formas em instâncias do arquivo da matriz.

Figura 9: : Organização do conjunto de dados: Quantidade de formas/distribuição das formas em instâncias (linhas).



5.3 O treinamento

A modelagem da rede neural foi realizada em linguagem *R* usando o pacote *Caret* (KUNH, 2008), versão 6.0-41, método '*nnet*'. O método *nnet* tem como base um *perceptron* multicamadas. Mais especificamente, o método *nnet* é especializado em redes neurais multicamadas de tipo *feedforward* para o aprendizado supervisionado. O treinamento em *nnet* utiliza o algoritmo *Backpropagation* tradicional.

O treinamento da rede para as simulações apresentadas neste capítulo utilizou uma técnica conhecida como validação cruzada (*cross-validation*). A validação cruzada é uma técnica útil em determinar, durante o treinamento, a capacidade de generalização do modelo. Por meio da validação cruzada, é possível selecionar o modelo e os parâmetros de modelagem que exibem melhor desempenho na classificação dos substantivos inanimados do português (presentes no *corpus*) em categorias de gênero. Selecionar o modelo com os melhores parâmetros de modelagem é importante para evitar situações de *overfitting* ou *underfitting*. Uma situação de *overfitting* ocorre quando um modelo apresenta uma arquitetura muito complexa para uma

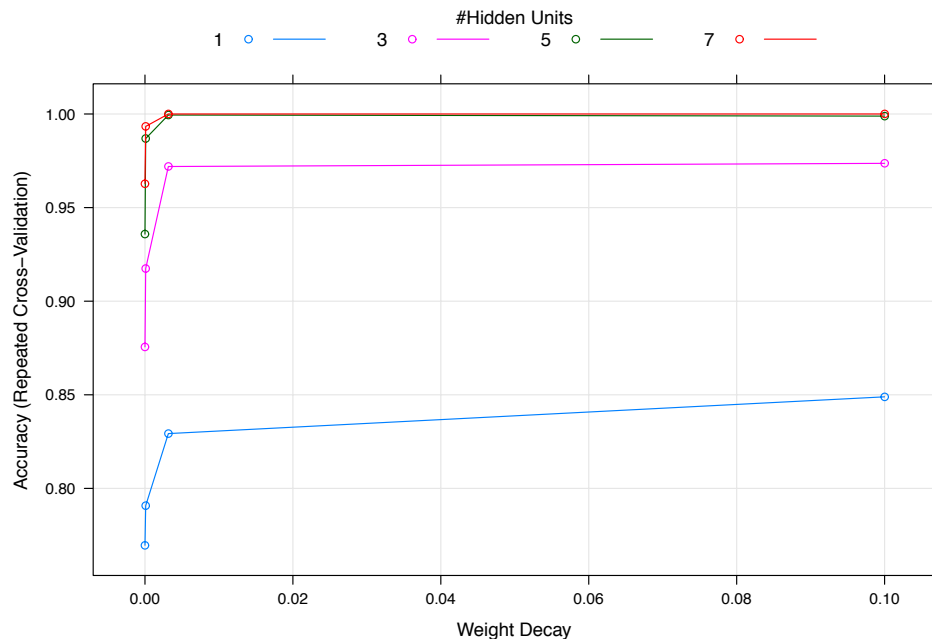
quantidade de informação pequena contida no banco de dados. *Underfitting* é a situação contrária, isto é, quando a arquitetura e os parâmetros selecionados para o modelo não dão conta da quantidade de informação contida no banco de dados, comprometendo a capacidade de generalização do modelo.

Para o treinamento da rede empregando a técnica de validação cruzada, foi preciso dividir o *corpus* em 2 subconjuntos: subconjunto de treino e subconjunto de teste. O subconjunto de treino, por sua vez, foi dividido em blocos: as 1783 instâncias do conjunto de dados separado para o treino ($\pm 75\%$ das 2386 instâncias) foram divididas em dez blocos (*ten-fold cross validation*). A validação cruzada consiste na modelagem de um dos dez blocos de dados. O modelo resultante é então testado (validado) em outro bloco de dados. Esse processo continua até que todos os blocos tenham sido modelados e testados. O desempenho final é baseado na acurácia das previsões dos modelos ao longo do processo de validação.

Para a modelagem empregada no presente estudo, o processo da validação cruzada foi repetido cinco vezes e programado para comparar quatro arquiteturas diferentes de modelos em 1000 épocas. O modelo final é o que apresentou melhor desempenho médio na validação cruzada. Figura 10, a seguir, ilustra a comparação do desempenho dos quatro modelos resultantes do treinamento empregando a técnica da validação cruzada. Os modelos diferenciam-se em relação aos seguintes parâmetros: nº de neurônios na camada intermediária (*hidden units*: 1, 3, 5 e 7) e taxa de *weight decay*¹⁵ (0.0 a 0.1).

¹⁵ *Weight decay* é uma forma de limitar a complexidade da rede neural por meio da multiplicação dos pesos por um fator menor que 1. Esse processo evita que os pesos se tornem muito grandes. A taxa de *weight decay* ‘penaliza’ pesos que se tornam muito grandes sem necessidade.

Figura 10: Comparação do desempenho de quatro modelos com diferentes parâmetros/complexidades.



Como se pode observar na Figura 10, a arquitetura do modelo ótimo final selecionado para as simulações apresentadas neste capítulo compreende nove neurônios na camada de entrada (*inputs*), representando o número máximo de grafemas dos substantivos selecionados para a modelagem, sete neurônios na única camada intermediária (para armazenamento e processamento dos padrões) e um neurônio na camada de saída, representando a resposta do modelo (gênero masculino ou feminino). A taxa de *weight decay* selecionada para o modelo final foi de 0.1. O modelo final apresenta uma acurácia de classificação estimada em 100%.

5.4 Simulação 1

Os resultados dos Experimento 1 e Experimento 2 do Capítulo 4 mostraram que os falantes do português são sensíveis às propriedades estruturais e à frequência de ocorrência dos itens lexicais. Em termos gerais, os participantes foram mais rápidos para concordar ou atribuir gênero a formas regulares e transparentes do que a formas opacas e irregulares, principalmente no que diz respeito a formas frequentes.

A Simulação 1 teve como objetivo verificar se o modelo selecionado com base em seu desempenho na validação cruzada, assim como os falantes do português, é sensível à frequência de ocorrência dos

itens. Para tanto, o modelo foi treinado com 75% das instâncias (1783 substantivos presentes no conjunto de dados) e testado nos 25% das instâncias restantes (603 instâncias). Vale lembrar que, apesar de várias estruturas presentes no conjunto de dados de treino apresentarem uma relação direta com as categorias de gênero, como formas regulares (terminação em *-a/-o*) e transparentes (terminação em *-agem, -or*), outras estruturas podem ser confundidoras, como as irregulares (terminação em *-ema* e *-ão*) bem como opacas (terminação em *-e*) que podem ser classificadas tanto como masculinas como femininas.

Relativamente às duas últimas categorias de formas acima citadas, é possível que o sistema leve em consideração a sua frequência de ocorrência no conjunto de dados para sua decisão na classificação. Vale lembrar que o arquivo de treino foi construído de maneira a conter somente substantivos de frequência média a alta. Assim sendo, minha hipótese é: quando o modelo não tiver a pista da terminação para a decisão de gênero, então a frequência de ocorrência das propriedades estruturais das formas como um todo (e não somente sua terminação) são consideradas pelo modelo para sua classificação.

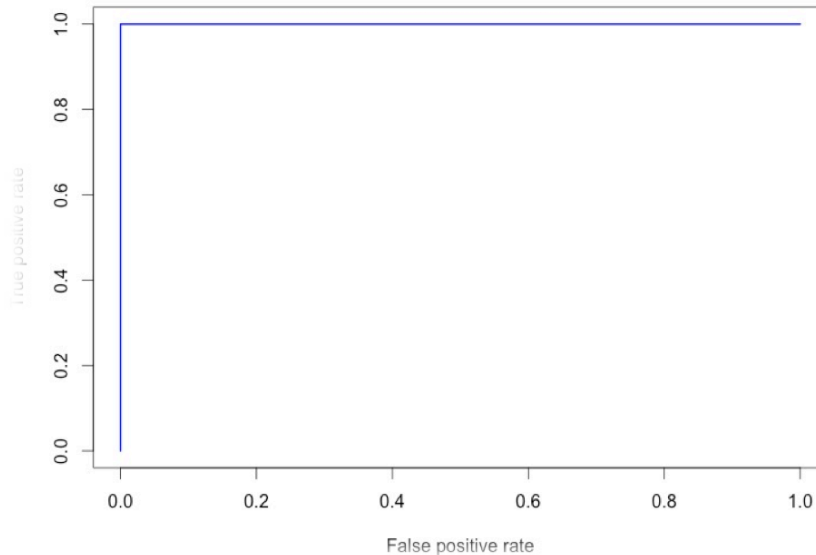
5.4.1 Resultados da Simulação 1

A matriz de confusão (Tabela 13), abaixo, contrasta a saída da rede com a saída desejada, e a Figura 11 apresenta a curva ROC (*Receiver Operating Characteristics*) que contrasta os verdadeiros positivos (acertos da rede) com os falso positivos (erros da rede).

Tabela 13: Matriz de confusão: classificação da rede vs. classificação real. Teste para 603 instâncias.

Porcentagem de dados para treinamento		Feminina	Masculina	Acurácia
75%	feminina	284	0	1
	masculina	0	319	

Figura 11: : Curva ROC. Simulação 1.



Conforme ilustram a matriz de confusão (Tabela 13) e a curva ROC (Figura 11), o modelo final selecionado foi capaz de classificar corretamente em categorias de gênero as 603 instâncias presentes no conjunto de dados separado para o teste. Esse resultado revela que os parâmetros selecionados para compor a arquitetura do modelo foram satisfatórios.

O desempenho de 100% do algoritmo para classificação de 25% das instâncias indica que a rede foi capaz de aprender de forma eficaz a relação entre estrutura e gênero e generalizar o conhecimento para formas não vistas. Vale lembrar que o algoritmo teve que lidar com formas ambíguas quanto ao gênero para classificação como, por exemplo, as formas irregulares (masculinas com terminação em *-a*, e femininas com terminação em *-o*) e formas opacas quanto ao gênero como substantivos femininos e masculinos com terminação em *-e*. Portanto, uma acurácia de 100% sugere que, mais que somente a terminação, a rede é capaz de estabelecer associações com base na frequência de ocorrência dos itens apresentados à rede na fase de treinamento. Nesse sentido, os resultados indicam que a rede, assim como os sujeitos, é sensível à frequência de ocorrência dos itens lexicais. Esse desempenho indica também que, embora o algoritmo tenha sido treinado com formas ambíguas quanto ao gênero (femininas e masculinas com terminação em *-e*, femininas e masculinas com

terminação em *-a* e *-o*), a rede neural foi capaz de aprender a relação entre forma e gênero com base não somente na relação entre estrutura e gênero, mas, também, com base na frequência de ocorrência dos itens, não havendo diferenças de desempenho de classificação entre formas regulares, irregulares, transparentes e opacas quanto ao gênero do português. Portanto, a hipótese formulada para esta simulação foi confirmada: quando o modelo não tiver a pista da terminação para a decisão de gênero, então a frequência de ocorrência das propriedades estruturais das formas como um todo são consideradas pelo modelo para sua classificação.

5.5 Simulação 2

Como visto, os resultados da Simulação 1 sugerem que a rede é capaz de estabelecer associações entre estrutura ortográfica de um substantivo e seu gênero com base na frequência de ocorrência dos itens, já que estruturas ambíguas quanto ao gênero foram apresentadas à rede na fase de treinamento e foram corretamente classificadas em categorias de gênero. Resta investigar, no entanto, em que medida a rede é sensível às propriedades estruturais das formas. A classificação da rede terá como base a estrutura dos substantivos como um todo ou apenas na terminação? Na Simulação 2, investigo essa questão apresentando à rede neural pseudopalavras para classificação em categoria de gênero.

O objetivo específico da Simulação 2 foi investigar se a rede é capaz de generalizar o conhecimento adquirido na fase de treino com substantivos para pseudopalavras. Adoto aqui um paradigma experimental semelhante ao do Experimento 3 (cf. Cap. 4). Nessa Simulação, usei o modelo final treinado com 75% das instâncias do conjunto em virtude da estimativa de sua alta capacidade de generalização. Para o teste, seis pseudopalavras foram construídas com base em palavras de terminação ambígua (terminação em *-e*) presentes no conjunto de dados usado no treinamento: *atividade*, *nome*, *filme*, *time*, *noite*. Apenas o grafema inicial desses substantivos foi trocado, resultando em: *etividade*, *bome*, *bilme*, *fome*, *foite*. Portanto, a rede deve ser capaz de classificar três substantivos como masculinos e dois como femininos. Assim, será possível verificar se, assim como observado com participantes humanos, a rede generaliza o conhecimento adquirido no treino para estruturas semelhantes sem ter como a base a somente a

terminação. Em acréscimo, será possível verificar se a rede é sensível à frequência de ocorrência de uma estrutura e sua relação com uma categoria de gênero. Dado que a rede foi apresentada à cinco substantivos com terminação ambígua (final *-e*), sendo que três delas são do gênero masculino, será possível verificar se os possíveis erros da rede têm relação com a força da representação do gênero masculino em oposição à representação mais fraca do gênero feminino.

Acredito que a rede será sensível à força da representação das propriedades estruturais das formas e sua relação com a categoria de gênero masculino em virtude da força da representação do gênero masculino para substantivos de terminação *-e*.

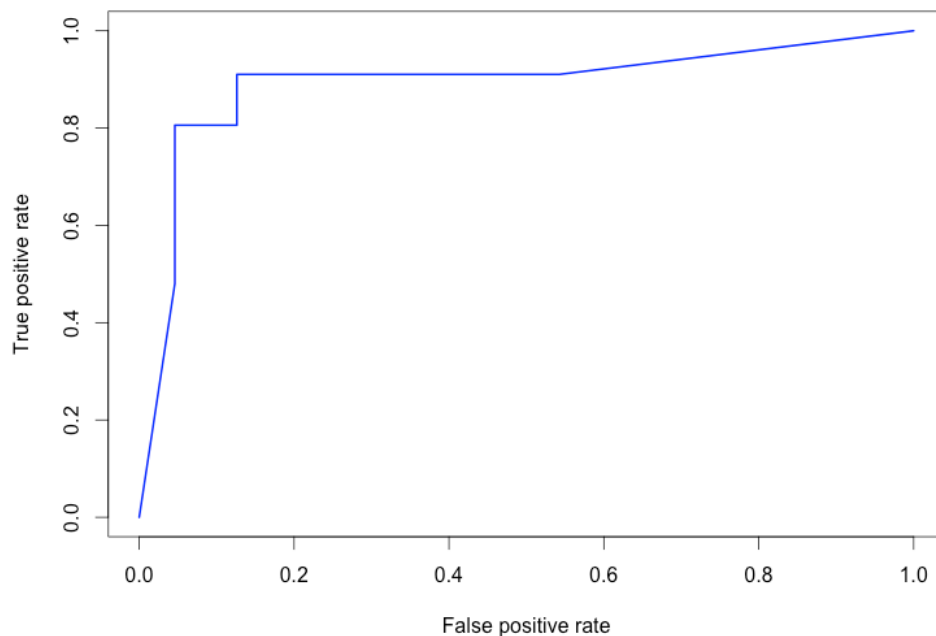
5.5.1 Resultados da Simulação 2

A matriz de confusão (Tabela 14), abaixo, contrasta a saída da rede com o saída desejada, e a Figura 12 apresenta a curva ROC (*Receiver Operating Characteristics*) que contrasta os verdadeiros positivos (acertos da rede) com os falso positivos (erros da rede).

Tabela 14: Matriz de confusão: classificação da rede *vs.* classificação desejada. Teste para 5 pseudopalavras.

Porcentagem de dados para treinamento		Feminina	Masculina	acurácia
75%	feminina	1	0	0.8
	masculina	1	3	

Figura 12: Curva ROC. Simulação 2



Conforme exhibe a matriz de confusão (Tabela 14), três substantivos masculinos foram classificados como masculinos. Portanto, como esperado, a rede foi capaz de generalizar o conhecimento adquirido na fase de treino para pseudopalavras não vistas pela rede. Tendo em vista que a terminação dos substantivos incluídos no treino era ambígua quanto ao seu gênero (femininos e masculinos com terminação em *-e*), para uma classificação precisa, foi necessário que a rede neural estabelecesse uma associação entre a estrutura da pseudopalavra com a estrutura ortográfica dos substantivos presentes no treinamento. Entretanto, um substantivo feminino foi classificado como masculino, resultando em uma acurácia de 80%. Esse erro de classificação parece estar relacionado à força da representação da terminação *-e* para o gênero masculino. No entanto, essa taxa de erro é pequena se levarmos em conta que a amostra (apenas 5 pseudopalavras) do teste é limitada. Contudo, os resultados sugerem que a rede é sensível às estruturas dos substantivos como um todo, um comportamento também observado pelos participantes do Experimento 3 independentemente da categoria de forma. Por outro lado, o resultado da saída da rede sugere, também, que a rede é sensível à força da representação de uma estrutura (terminação *-e*) e a relação com as

categorias de gênero (gênero masculino). A Simulação 3, apresentada a seguir, foi desenhada para buscar mais dados para confirmar essa hipótese.

5.6 Simulação 3

Conforme sugere a Simulação 2, dada uma estrutura ambígua que pode ser classificada tanto como feminina quanto masculina, a rede neural leva em consideração todas as pistas fornecidas pela estrutura (não somente sua terminação) para sua decisão de classificação, conforme o comportamento observado na tarefa de atribuição de gênero do Experimento 3 no Capítulo 4.

Na Simulação 3, abordo a mesma questão da Simulação 2. No entanto, o problema apresentado à rede neural na Simulação 3 diferencia-se do problema da Simulação 2 em um ponto crucial: Assim como na Simulação 2, na Simulação 3, a rede deve ser capaz de classificar pseudopalavras de terminação ambígua em categorias de gênero. No entanto, a ambiguidade para a classificação está presente apenas no grafema final *-a*, já que três das pseudopalavras apresentadas à rede neural para teste são modificações de substantivos regulares (femininos terminados em *-a*), e outras três são modificações de substantivos irregulares (masculinos terminados em *-ema*). Como já mencionado, a categoria das regulares compreende mais de 50% das instâncias do arquivo de treino (22 substantivos de 40). Nesse sentido, a frequência do padrão da terminação em *-a* para o gênero feminino é maior que o padrão da terminação em *-ema* para o gênero masculino. Portanto, é possível que a rede ‘regularize’ as pseudopalavras irregulares em função da frequência superior de ocorrência do padrão terminação em *-a* para o gênero feminino.

No Experimento 3 do Capítulo 4, os resultados mostraram que os participantes ‘regularizaram’ as pseudopalavras de difícil associação (pseudopalavras cuja pista para dedução do gênero estava apenas no último grafema). Essa ‘regularização’ das formas irregulares quanto ao gênero manifestou-se tanto no tempo de reação aos estímulos quanto na acurácia: os participantes exibiram um tempo de reação menor (927ms), aproximando-se do tempo de reação médio de pseudopalavras regulares (905ms) e atribuíram a essas, em sua maioria, o gênero feminino (71,4% das possibilidades de resposta). Por outro

lado, para as pseudopalavras irregulares terminadas em *-ema* de fácil associação (pseudopalavras que simulavam pares mínimos com substantivos pré-existentes no léxico mental dos falantes), essa taxa de ‘regularização’ diminuiu na classificação em categorias de gênero (61,9%), bem como aumentou o tempo médio de reação (979ms), sugerindo que um processo de associativo foi estabelecido entre substantivos pré-existentes no léxico mental dos falantes e pseudopalavras (cf. Cap. 4, seção 4.4.3).

A intenção da Simulação 3 é investigar em que medida a rede é sensível às semelhanças estruturais existente entre as pseudopalavras apresentadas para teste e os substantivos apresentados à rede na fase de treino. Além disso, investigo, também, em que medida a rede é sensível à frequência de ocorrência de um padrão em oposição à infrequência de outro levando em conta que esses padrões podem atuar como um fator confundidor para a rede em virtude da ambiguidade do grafema final *-a* presente tanto nas regulares quanto nas irregulares.

Se a rede classificar as pseudopalavras terminadas em *-ema* como femininas, então, assim como os participantes humanos, a rede neural estará ‘regularizando’ as irregulares em virtude da força associativa da terminação *-a* ser maior (mais frequente) para o gênero feminino que a força associativa da terminação *-ema* é para o gênero masculino. Se, por outro lado, a rede classificar pseudopalavras com terminação em *-ema* como masculinas e pseudopalavras terminadas em *-a* como femininas, então é possível concluir que a saída da rede está se baseado na regra da terminação. Se assim for, a rede será capaz de distinguir 3 pseudopalavras femininas de 3 pseudopalavras masculinas.

O modelo ótimo final selecionado para as Simulação 1 e Simulação 2 foi também usado na Simulação 3. Para essa Simulação, seis pseudopalavras foram construídas diferenciando-se dos substantivos reais do português apresentados à rede em apenas um grafema. Os substantivos femininos com terminação em *-a* foram: *cultura*, *economia*, *pesquisa*. Os três substantivos masculinos com terminação em *-ema* foram: *cinema*, *tema* e *sistema*. As pseudopalavras construídas com base nessas formas foram: *culcura*, *etonomia*, *mesquisa*, *cilema*, *fema* e *mistema*. Minha hipótese é: a rede será sensível à força da representação das propriedades estruturais das formas e sua relação com a categoria de gênero feminino em virtude da força da representação do gênero

feminino para substantivos de terminação *-a*. Portanto, é possível que a rede ‘regularize’ (algumas) formas irregulares (terminação em *-ema*).

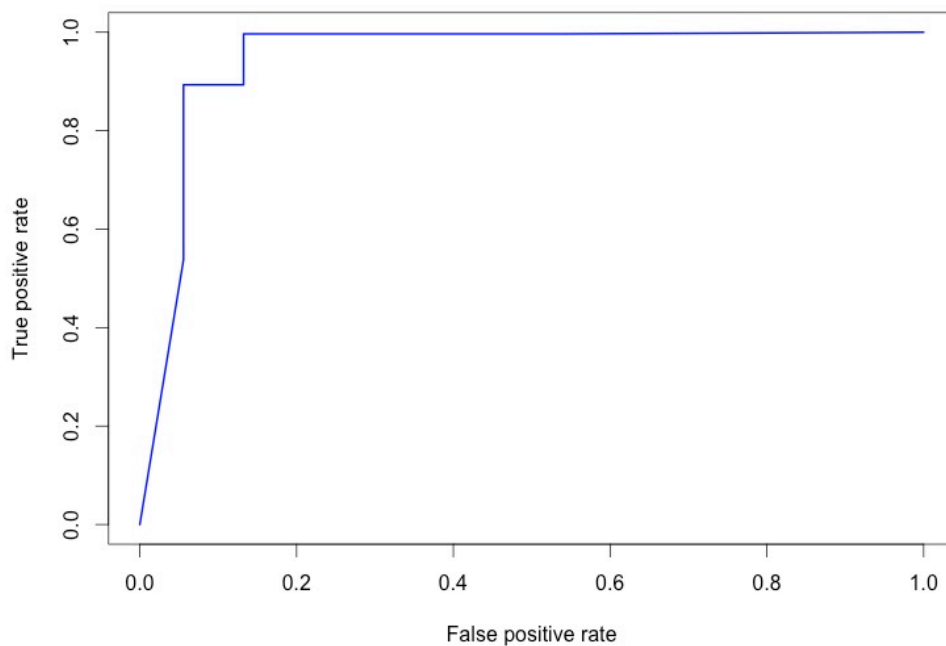
5.6.1 Resultados Simulação 3

A matriz de confusão (Tabela 15), a seguir, contrasta a saída da rede com a saída desejada, e a Figura 13 apresenta a curva ROC (*Receiver Operating Characteristics*) que contrasta os verdadeiros positivos (acertos da rede) com os falso positivos (erros da rede).

Tabela 15: Matriz de confusão: Classificação da rede vs. Classificação desejada. Teste para 6 pseudopalavras.

Porcentagem de dados para treinamento		feminina	masculina	acurácia
75%	feminina	3	1	0.83
	masculina	0	2	

Figura 13: Curva ROC. Simulação 3.



Conforme exhibe a matriz de confusão (Tabela 15), a saída do modelo para a classificação de seis pseudopalavras em categorias de gênero apresenta uma acurácia de 83%. As três pseudopalavras com terminação em *-a* foram classificadas como femininas. Duas pseudopalavras com terminação em *-ema* (típica do gênero masculino) foram classificadas como masculinas e uma classificada como feminina. Esse resultado sugere, por um lado, que o modelo estabelece associações entre a estrutura das formas como um todo e não somente o grafema final e, de outro, que o modelo ‘regularizou’ uma forma irregular quanto ao gênero. Essa porcentagem de erro, apesar de pequena, indica que a força associativa da terminação *-a* venceu sobre a força associativa da terminação *-ema* em virtude da alta frequência do padrão *-a* e sua relação com o gênero feminino.

Como vimos na descrição do *corpus*, 11 substantivos de alta frequência terminados em *-a* foram associados ao gênero feminino, ao passo que apenas 3 substantivos terminados em *-ema* foram associados ao gênero masculino. Portanto, a força associativa do grafema final *-a* está mais frequentemente associado ao gênero feminino que a sequência *-ema* ao gênero masculino.

Esses resultados aproximam-se tanto da Simulação 2 quanto dos achados do Experimento 3 do Capítulo 4: tanto o modelo quanto falantes do português regularizam pseudopalavras terminadas em *-ema*, apesar da porcentagem de regularização dos falantes do português ter sido muito maior que a do modelo (61,9% e 71,4% contra 17%). Esses resultados indicam que a rede é sensível às estruturas dos substantivos como um todo e sua relação com uma categoria de gênero.

Com base nos resultados da Simulação 2 e Simulação 3 é possível concluir que a rede é sensível à força da representação de uma estrutura (nesse caso a força da terminação *-a*) e sua relação com a categoria de gênero feminino. Além disso, os dados mostram que a rede considera todos os traços compartilhados entre as estruturas para sua classificação, não somente a pista da terminação.

5.7 Discussão geral do capítulo

Neste capítulo, apresentei três simulações computacionais usando redes neurais artificiais que tiveram como objetivo modelar o processamento do gênero gramatical do português.

Na fase de treinamento da rede, a técnica da validação cruzada foi empregada para selecionar o modelo que apresenta parâmetros adequados e um desempenho satisfatório para a classificação de substantivos do português e pseudopalavras em categorias de gênero. O modelo ótimo selecionado foi composto por três camadas de neurônios: nove neurônios na camada de entrada, sete neurônios na camada intermediária e um neurônio na camada de saída. Como *inputs*, o modelo recebeu a representação da sequência de grafemas de cada substantivo, simulando uma tarefa visual de atribuição de gênero. Como *outputs*, o modelo produziu as categorias de gênero, a saber: ‘F’ para o gênero feminino, ‘M’ para o gênero masculino. O modelo selecionado apresentou um acurácia estimada em 100% .

Na Simulação 1, como previsto, o modelo apresentou uma performance de 100% na classificação de 603 palavras em categorias de gênero. Esse resultado evidencia a eficiência do modelo em aprender os padrões estruturais dos substantivos a ele apresentados e relacioná-los com as categorias de gênero. Mesmo em se tratando de estruturas ambíguas quanto ao gênero, como substantivos com terminação em *-e*, o modelo foi capaz de classificá-las em categorias de gênero corretamente. Esse resultado sugere um efeito da frequência dessas formas ambíguas, já que, apesar de ausência de pistas estruturais para a classificação em categorias de gênero, o modelo foi capaz de reconhecer as propriedades estruturais dessas formas como um todo com base em sua frequência de entrada no conjunto de dados de treino.

A Simulação 2 teve como objetivo investigar em que medida a rede neural é sensível à frequência de uma propriedade estrutural e sua relação com uma das categorias de gênero. Nessa simulação, a rede foi apresentada a pseudopalavras com terminação em *-e* as quais se distinguiam de substantivos terminados em *-e* presentes no treinamento em apenas um grafema. A rede apresentou uma acurácia de 80% de classificação, com apenas uma classificação errada em cinco possibilidades: as três pseudopalavras que compartilhavam traços estruturais com palavras masculinas foram classificadas como masculinas; das duas pseudopalavras que compartilhavam traços estruturais com palavras femininas, uma foi classificada como feminina e, a outra, como masculina. Esse erro de classificação sugere o efeito da frequência da estrutura e sua relação com o gênero masculino, visto que a maior parte das formas com terminação em *-e* apresentadas à rede no treinamento eram formas masculinas. Esse mesmo efeito, isto é, o efeito da força de uma estrutura e sua relação com uma categoria de gênero foi observado na Simulação 3.

Na Simulação 3, a rede foi apresentada a estruturas terminadas em *-a*, e masculinas terminadas em *-ema*. Portanto, um problema de ambiguidade de classificação de gênero foi apresentado à rede, já que a terminação *-ema* é tipicamente masculina, mas termina em *-a*.

Na Simulação 3, o modelo de rede neural apresentou uma acurácia de 83%, aproximando-se da acurácia da Simulação 2: três pseudopalavras femininas foram classificadas como femininas; de três pseudopalavras masculinas (terminação em *-ema*), apenas duas foram classificadas como masculinas e uma como feminina. Apesar do poder de generalização da rede ter sido alto, em um caso, a força da representação da terminação *-a* para o gênero feminino venceu na competição de pesos com a sequência de grafemas finais *-ema* típico do gênero masculino. Juntos, esses resultados sugerem que rede é sensível à aos traços compartilhados entre as estruturas com um todo e não apenas à pista presente no grafema final para classificação em gênero. Sugerem, ainda, que a rede é sensível a força da representação de uma estrutura (em virtude de sua frequência de ocorrência) e a relação dessa estrutura com uma categoria de gênero.

Os achados das simulações aqui empregadas contribuem para o debate existente entre modelos de via dual e modelos de via única. Quando analisados em conjunto, os resultados mostram que um simples modelo conexionista foi capaz de aprender as propriedades estruturais dos substantivos do português regulares/irregulares/opacas e transparentes quanto ao gênero e eficientemente relacioná-las com a categorias de gênero. Em acréscimo, as saídas do modelo para a classificação de pseudopalavras em categoria de gênero mostraram que, conforme postula o *Network model*, a força de representação de uma estrutura depende de sua frequência de ocorrência. Assim como demonstrado no Experimento 3 do Capítulo 4, a força da representação da terminação *-a* para o gênero feminino venceu sobre a sequência de grafemas *-ema* para o gênero masculino em virtude de sua maior quantidade de ocorrência no conjunto de dados separados para o treinamento. De forma análoga, o modelo ‘regularizou’ parte das formas irregulares. Semelhantemente, a força da representação do grafema final *-e* para o gênero masculino venceu a competição de pesos sobre o gênero feminino em virtude da terminação *-e* ter sido mais fortemente associada ao gênero masculino que ao gênero feminino durante o treinamento.

Vale lembrar que os resultados da saída da rede para as Simulações 2 e 3 empreendidas assemelham-se aos resultados encontrados nos experimento psicolinguísticos com falantes do

português. Assim como a rede, falantes do português regularizaram parte das formas irregulares (terminação em *-ema*) no Experimento 3. No entanto, essa taxa de regularização decresceu à medida que os traços compartilhados com palavras pré-existentes em seu léxico mental aumentaram. Portanto, apesar da força de representação de *-a* ter vencido, em alguns casos, tanto no experimento psicolinguístico quanto na classificação da rede neural, os traços compartilhados entre as pseudopalavras e substantivos cumpriram um papel crucial para a decisão de gênero tanto no que diz respeito à classificação realizada pelos falantes do português, quanto no que diz respeito à classificação realizada pela rede neural.

Para terminar, é importante esclarecer que a modelagem conexionista apresentada neste capítulo buscou cobrir toda uma variedade de categorias de formas e a relação dessas formas com as categorias de gênero do português, bem como introduziu no treinamento substantivos de comprimentos diversos (entre três e nove grafemas) e de complexidade estrutural diversa: entre uma e quatro sílabas, com ou sem acentuação. Além disso, o modelo foi treinado com uma proporção de frequência de entrada de substantivos que se aproximada à frequência de ocorrência desses substantivos na língua portuguesa. Portanto, as generalizações feitas pelo modelo aqui apresentado são baseadas em análises estatísticas que consideraram a frequência das formas e seus padrões ortográficos sequenciais como um todo e não em codificação de regras.

CAPÍTULO 6

6 OS CORRELATOS NEUROFISIOLÓGICOS DO PROCESSAMENTO DA CONCORDÂNCIA DE GÊNERO DO PORTUGUÊS

6.1 Introdução

Estudos que versam sobre o processamento do gênero gramatical ou, mais amplamente, que versam sobre o processamento e representação mental do gênero gramatical de línguas do mundo adotam, tradicionalmente, o paradigma comportamental de experimentação o qual baseia-se, essencialmente, em tarefas de decisão metalinguística de gênero. Em linhas gerais, nessas tarefas, os sujeitos participantes dos experimentos são requisitados a indicar o gênero de uma palavra, apresentada visualmente ou auditivamente, pressionando botões de uma caixa de resposta associados a uma categoria de gênero. Os resultados desses estudos são interpretados com base na análise do tempo de reação aos estímulos e acurácia.

Apesar de essa abordagem experimental ter se tornado um paradigma profícuo para o entendimento da aquisição e processamento do gênero, é possível encontrar na literatura sobre o tópico resultados controversos os quais têm suscitado discussões em torno da dualidade *versus* a singularidade das rotas de processamento do gênero gramatical.

Embora exista uma convergência de resultados no que se refere ao tempo de reação na comparação do processamento de palavras transparentes, opacas e regulares e irregulares (em grande parte do estudos, os sujeitos são mais rápidos para atribuir gênero a palavras transparentes e regulares quanto ao gênero que a opacas e irregulares quanto ao gênero) (TAFT; MEUNIER, 1998; DESROCHERS *et al.*, 1989; AFONSO *et al.*, 2013; GOLLAN; FROST, 2001; BATES *et al.*, 1995), quando fatores não estruturais são manipulados, como a (in)frequência de ocorrência dos itens e a quantidade de vizinhos ortográficos de uma palavra, os resultados divergem e, por consequência, a interpretação do fenômeno por parte dos pesquisadores. Por exemplo, enquanto Gollan e Frost (2001) assumem, com base no contraste do tempo de reação a formas marcadas e não marcadas quanto ao gênero equivalentes em frequência, que o processamento do gênero gramatical do hebraico é governado por uma dupla rota - uma de caráter automático/implícito para o processamento das estruturas fonológicas regulares quanto ao gênero, e uma memória de caráter explícito/associativo responsável pelo armazenamento do gênero das formas, resultados de estudos anteriores, como os de Taft e Meunier

(1998) e Desrochers (1989), que manipularam a frequência dos itens, mostraram um efeito individual da frequência, sugerindo a extração de todos os itens da memória para o processamento de seu gênero independentemente de suas formas. Por outro lado, vários estudos usando técnicas de neuroimagem para o escaneamento cerebral dos sujeitos na execução de uma tarefa de decisão de gênero mostraram diferenças no padrão de ativação na decisão de gênero de palavras regulares, irregulares, transparentes e opacas. Entre essas investigações, destacam-se os estudos que abordaram a questão das rotas do processamento do gênero gramatical nas línguas românicas.

Hernandez *et al.* (2004), por exemplo, investigaram as bases neurais do processamento do gênero gramatical regular (femininas com terminação em *-a*, e masculinas com terminação em *-o*) e irregular (femininas com terminação em *-o*, e masculinas com terminação em *-a*) do espanhol usando ressonância magnética funcional (fMRI). Os resultados mostraram ativação das zonas BA44 e BA45 (áreas de Broadman) localizadas na região de Broca (giro inferior frontal do córtex). No entanto, os resultados mostram que houve uma maior ativação na região BA45 para a decisão de gênero de palavras irregulares. Esse mesmo resultado foi replicado em um estudo posterior. Hernandez *et al.* (2007) investigaram se há diferenças na maneira como falantes do espanhol como segunda língua precoces e tardios processam o gênero de formas regulares e irregulares. Novamente, os resultados mostraram forte ativação das regiões BA44 e BA45. A região BA45 foi mais ativada para formas irregulares em ambos os grupos de falantes do espanhol.

Um estudo similar aos anteriormente mencionados foi conduzido por Padovanni *et al.* (2005) que investigaram se, no italiano, há um aumento gradual na ativação na decisão de gênero de formas regulares/transparentes, opacas e irregulares. Os resultados mostraram um aumento gradual na ativação na BA44: de transparente para opacas e de opacas para irregulares. Esse aumento na ativação foi acompanhado por um aumento gradual no tempo de reação na decisão do gênero de palavras transparentes, opacas e irregulares. Na interpretação dos autores, esses diferentes padrões de ativação sugerem que formas regulares são processados por mecanismos neurocognitivos distintos e que a região de Broca, mais especificamente as zonas BA 44 e BA45, cumpre papel essencial no processamento do gênero.

Cabe ressaltar, contudo, que tanto os estudos que empregaram o escaneamento cerebral quanto os estudos comportamentais acima elencados usaram uma tarefa de decisão de gênero. Essa tarefa apresenta

desvantagens para uma boa interpretação dos dados em virtude de sua artificialidade (cf. HEIM, 2008). Ao usarem a língua, falantes nativos não precisam decidir o gênero das palavras, isto é, não precisam usar seu conhecimento metalinguístico sobre o gênero da palavra para corretamente realizar a concordância. Assim, uma questão importante que deve ser considerada é: Em que medida os dados de neuroimagem (principalmente usando ressonância magnética) e dados provenientes dos estudos comportamentais acima descritos podem ser generalizados para o processamento do gênero em uma situação de fala normal que não requer a decisão explícita de gênero?

Uma técnica de neuroimagem não invasiva que permite analisar o processamento da língua em tempo real e em situações normais de uso da língua, como a simples leitura ou audição de frases, é a extração dos potenciais relacionados a eventos (ERPs - *event related potentials*) (cf. KAAN, 2007) por meio de um eletroencefalograma (EEG). Um estudo de ERP é vantajoso em relação a outras técnicas de neuroimagem porque não requer que os sujeitos executem uma tarefa, a menos que a extração dos ERPs relacionados a execução de uma tarefa seja objetivo do estudo.

Kutas e Van Peten (1994) explicam que o EEG tira vantagem do fato de processos mentais de transmissão de informação no cérebro envolver fluxo de íons. Esse fluxo de íons gera um campo de voltagem¹⁶ ao redor de cada neurônio ativo. A soma dos campos elétricos produzidos por neurônios vizinhos pode ser detectada por eletrodos. Mais especificamente, os eletrodos captam as flutuações de voltagens produzidas por uma dada população de neurônios. Portanto, o EEG reflete as voltagens produzidas pela atividade elétrica cerebral para uma atividade ao longo do tempo. Portanto, um estudo de EEG consiste na gravação da atividade elétrica (detectada por eletrodos) de um sujeito durante a execução de uma tarefa ou a simples leitura ou audição de um estímulo de interesse para o pesquisador.

Com o avanço na tecnologia de computadores, foi possível extrair a média dessas atividades neuronais com o propósito de obter uma estimativa da atividade em relação a um determinado ponto de um evento externo. Um ponto em um evento externo seria, por exemplo, a leitura ou audição de uma determinada palavra em uma frase. Para obter a estimativa da atividade de um ponto em um evento externo é preciso

¹⁶ A voltagem pode ser definida como a diferença do potencial elétrico entre dois lugares, como entre dois pontos do escalpo.

trancar esse ponto no tempo, isto é, trancar a gravação da atividade elétrica (flutuações de voltagens) a partir do momento em que um estímulo é apresentado (*onset*). A média dos potenciais gerados pela repetição de estímulos de mesmo tipo trancados no tempo é então calculada. Essa média dos potenciais gerados com a repetição de estímulos é conhecida como potenciais relacionados a eventos ou ERPs (do inglês: *Event related potentials*). O cálculo da média dos potenciais gerados por vários estímulos de mesmo tipo realçam os potenciais associados ao *onset* do estímulo de interesse para o pesquisador e, ao mesmo tempo, reduz os potenciais que não estão relacionados ao *onset* do estímulo de interesse.

É importante salientar que as ondas de ERPs trancadas no tempo (*time-locked*) podem ser de picos positivos ou negativos. Geralmente, os gráficos que exibem as ondas de ERPs mostram ondas negativas para cima e positivas para baixo. Os picos de um ERP são nomeados de acordo com sua polaridade e sua latência em milissegundos em relação ao *onset* do estímulo. Por exemplo, se um pico é positivo e surge após 300 milissegundos depois da apresentação de um estímulo, então, esse ERP é denominado P300.

Para a análise do processamento da linguagem, alguns componentes de ERP são importantes, principalmente, os componentes N400, P600 e (E)LANs. A seguir, apresento uma breve descrição de cada um desses componentes:

N400. Em um estudo seminal, Kutas e Hillyard (1980) reportaram uma onda negativa com pico entre 300ms e 400ms relacionada a uma incongruência semântica. Por exemplo, na frase **hoje eu tomei café com taxi* há apenas incongruência semântica, já que a frase é sintaticamente gramatical. Tipicamente, o pico máximo do N400 é atingido nas regiões centro-parietais do escalpo após a apresentação do estímulo incongruente (cf. KUTAS; FEDERMEIER, 2010). As incongruências semânticas, também denominadas na literatura como violações semânticas, que geram esse componente podem ser de tipo linguístico (KUTAS; HILLYARD, 1980; FRIEDERICI; PFEIFER; HAHNE, 1993) e não linguístico (HAGOORT, 2011).

Em relação às incongruências linguísticas, Kutas e Federmeier (2010) explicam que a latência e a amplitude do N400 dependem do tipo de violação semântica: quanto mais inesperada a palavra no contexto semântico em que está inserida, maior a amplitude da negatividade.

Assim, esse componente está altamente relacionado a expectativa de encaixe da palavra no contexto semântico da sentença.

O componente N400 tem sido observado, também, no processamento morfológico. Em um estudo seminal, Münte *et al.* (1999) usaram uma tarefa de *priming* com o objetivo de investigar diferenças entre os processamentos da morfologia flexional verbal regular e irregular do passado do inglês. Nesse estudo, 19 sujeitos leram a forma infinitiva de verbos regulares e irregulares as quais eram precedidas (*primed*) ou não (*unprimed*) por suas formas flexionadas no passado. Os resultados mostram uma redução do N400 apenas no processamento de verbos regulares do passado do inglês que eram precedidos por sua forma flexionada (condição *primed*). O mesmo efeito não foi encontrado no processamento de itens irregulares na condição *primed*. Na interpretação dos resultados, os autores concluem que apenas verbos regulares são decompostos em sua forma durante o processamento.

Coulson *et al.* (1998) esclarecem que menores amplitudes do componente N400 têm sido interpretadas como reflexo de uma redução na dificuldade de integração semântica das sentenças.

P600. No que se refere ao processamento de tipo sintático e morfossintático, estudos eletrofisiológicos têm documentado o aparecimento de uma onda de valência positiva e latência entre 500ms a 900ms (cf. FRIEDERICI; HAHNE; SADDY, 2002) para violações de tipo sintática/morfossintática na língua materna, principalmente, para estímulos com incongruência na concordância. Essa positividade foi encontrada para anomalias na concordância de número, de gênero (DOWNES *et al.* 2011; BANÓN, FIORENTINO; GABRIELE, 2012) e na concordância verbal, bem como na estrutura sintática de sentenças (BOWDEN, 2007). Essa onda, denominada P600, atinge seu pico máximo nas regiões centro-parietais, apesar de sua larga destruição no escalpo (cf. KAAN; SWAAB, 2003).

Segundo Hahne e Friederici (1999), o componente P600 está associado a uma série de operações sintáticas e morfossintáticas além da re-análise de estruturas sintáticas violadas. Alguns autores (BARBER; CARRERAS, 2005; MOLINARO; VESPIGNANI 2008) apontam para a natureza bifásica do componente. De acordo com essa visão, a primeira fase ocorre entre 500 e 700ms e é associada a operações de ordem sintática como o processamento da ordem das sentenças e desambiguidade estrutural. A segunda fase, por seu turno, está associada a operações de re-análise do material linguístico, ou seja, uma fase associada a operações de ordem morfossintática em que os elementos são analisados em relação a outros elementos, como processos de

concordância de número, gênero, pessoa. (cf. KAAN; SWAAB, 2003). Uma segunda perspectiva (COULSON; KING; KUTAS, 1998; OSTERHOUT *et al.*, 1996) sugere, no entanto, que o P600 é interpretado como uma manifestação do P300, um componente que apresenta uma topografia semelhante a do P600. Coulson, King e Kutas (1998) afirmam que o P600 está associado a apenas ao P3b, um subcomponente da família do P300 que atinge sua amplitude máxima nas regiões centro-parietais. O P300 e seus subcomponentes estão associados a processamentos extra-linguísticos, por exemplo, processos que envolvem a percepção de aspectos subjetivos dos estímulos como a sua relevância na tarefa, saliência e probabilidade (cf. COULSON, KING, KUTAS, 1998).

Alguns estudos mostram que o P600 é gerado, preferencialmente, em tarefas de julgamento de gramaticalidade de sentenças (BANÓN; FIORENTINO; GABRIELE, 2012; BARBER; CARREIRAS, 2003). No entanto, outros estudos (DOWENS *et al.*, 2011; HAGOORT, 1993; MUNTE *et al.* 1994) mostram que o componente também é gerado em tarefas que não exigem o julgamento de sentenças, mas apenas a leitura de sentenças contendo uma incongruência de tipo morfossintático. Portanto, esses estudos sugerem que o P600 está associado também a operações de cunho implícito. Vale ressaltar que o P600 é evocado, na maioria dos estudos, por violações relacionadas à concordância independentemente do tipo, da língua sob investigação e do tipo de tarefa usada no experimento.

No estudo de Hagoort *et al.* (2003), o P600 apareceu em uma tarefa de leitura silenciosa de sentenças do holandês violadas e não violadas em gênero e número. Para o inglês, o estudo de Coulson, King e Kutas (1998) reportou o aparecimento do P600 na leitura de 200 sentenças com ou sem erros de concordância entre sujeito e verbo. No estudo de Barber e Carreras (2005), o P600 aparece também na leitura de sentenças em espanhol violadas na concordância de gênero.

Friederici (2002) ressalta que o P600 está associado não só ao processamento da re-análise de uma estrutura, mas, também, ao processamento de reparos de uma estrutura que apresenta uma violação. Esse tipo de processamento, geralmente, apresenta uma distribuição centro-parietal. No entanto, de acordo com outros autores, P600 apresenta uma distribuição mais posterior (cf. KAAN, 2007).

(E)LANs. Além dos P600, um outro componente relacionado ao processamento sintático e morfossintático é conhecido por LAN: *left anterior negativity*. Como o próprio nome sugere, LAN é uma negatividade mais proeminente na porção anterior do hemisfério

esquerdo. É possível encontrar na literatura dois tipos diferentes de LANs: ELANs e LANs (cf. KAAAN, 2007). A negatividade denominada ELAN (*early left anterior negativity*) ocorre, tipicamente, entre 100ms e 200ms após o *onset* do estímulo de interesse. O componente LAN, por sua vez, apresenta uma latência mais tardia com pico entre 300 e 400 ms após *onset* do estímulo.

Em linhas gerais, o componente ELAN está associado a processamentos de tipo automático, como o processamento da estrutura de um sintagma. Já o componente LAN, é tipicamente associado a dificuldades no processamento morfossintático, como dificuldades no processamento da concordância. Diferentemente do P600 que reflete processos tardios de re-análise e reparo de uma violação morfossintática, LANs refletem os primeiros estágios de detecção de uma anomalia na concordância (cf. KAAAN, 2007; FRIEDERICI; HAHNE; SADDY, 2002). Ullman (2001, 2004) assume que a LANs refletem processos típicos da memória procedural, como a construção da estrutura gramatical, tanto no que se refere aos processos que envolvem morfologia e sintaxe. Portanto, sob essa perspectiva, LANs refletem processos automáticos governados por regras.

Neste capítulo, dou prosseguimento ao debate que se abriu em torno do processamento do gênero gramatical de nomes substantivos inanimados. Nos capítulos anteriores, três estudos comportamentais foram conduzidos, assim como três simulações computacionais. Esses experimentos tiveram como objetivo geral investigar se o gênero gramatical do português é governado por uma única ou uma dupla rota de processamento. Os resultados dos experimentos apontam para um processamento do gênero de tipo associativo. No entanto, a interpretação dos dados está baseada em estudos que adotam um paradigma experimental que não reflete, conforme já mencionado, situações reais de uso da língua.

Em virtude de os estudos de ERPs permitirem que dados do processamento de situações reais de uso da língua sejam acessados, já que não demandam a execução de uma tarefa, essa técnica foi adotada no experimento apresentado neste capítulo. Busco, com este estudo, complementar a série de experimentos com foco no gênero gramatical do português a fim de buscar mais dados para o um melhor entendimento dos mecanismos que subjazem ao processamento desta categoria gramatical. Acredito que a análise dos ERPs ajudará a delinear os diferentes subprocessos envolvidos durante o uso do gênero gramatical em situações reais de uso da língua portuguesa.

6.2 O presente estudo

O objetivo do presente estudo é investigar se, assim como preconizam os modelos de via dual, o processamento do gênero de substantivos inanimados do português é governado por uma dupla rota, ou se, conforme preconizam os modelos unitários de processamento, é governado por um único sistema de memória associativa. Uma forma de investigar as rotas do processamento do gênero é dirigir nosso olhar para a relação existente entre forma e gênero.

Conforme apresentando no Capítulo 2 desta tese, no português, a terminação *-a* está em estreita relação com o gênero feminino, e a terminação *-o* em estreita relação com o gênero masculino, ao passo que as demais não são tão informativas quanto ao seu gênero, apesar de outras terminações do português serem também típicas do gênero feminino (*-agem*, *-ade*) e do gênero masculino (*-or*, *-ume*, *-ema*). Para formar o gênero feminino no português, *-a* comporta como a vogal *default*. Semelhantemente, a vogal átona final *-o* é a vogal *default* para a formação do gênero masculino (cf. NAME, 2002; NASCIMENTO, 2006; VILLALVA, 2000, CORRÊA, 2001). Os resultados dos estudos anteriormente apresentados neste trabalho (cf. Cap. 4) mostraram que os falantes nativos do português foram mais rápidos na execução das tarefas de concordância ou atribuição de gênero a formas regulares (femininas com final *-a*, e masculinas com final *-o*) que para as demais formas (transparentes, opacas e irregulares). Os resultados desses experimentos estão em consonância com os resultados de estudos que versam sobre o processamento do gênero gramatical em outras línguas, principalmente, estudos com foco no gênero gramatical de línguas românicas (AFONSO *et al.*, 2013, PADOVANI, 2002; PADOVANI; CACCIARI, 2003). Portanto, é possível que formas regulares quanto ao gênero sejam processadas por um mecanismo distinto do mecanismo responsável pelo processamento das demais formas.

Como vimos no referencial teórico desta tese (cf. Cap. 3), os modelos de via dual assumem que o processamento de formas regulares, altamente produtivas em uma língua, distigue-se do processamento de formas irregulares, mesmo daquelas que apresentam determinadas subregularidades. É possível, também, que o gênero de estruturas regulares, irregulares, transparentes e opacas seja processado pelos mesmos mecanismos neurocognitivos. Nesse caso, diferenças entre o processamento de formas regulares e o processamento das demais formas (transparentes, opacas e irregulares) seja fruto da força de representação das propriedades fonológicas dos itens (principalmente

terminação) e sua relação com o seu gênero gramatical.

No presente estudo, proponho como forma de investigar se itens regulares quanto gênero são processados pelos mesmos ou diferentes mecanismos neurocognitivos das demais formas, a extração e análise dos ERPs suscitados por ambas as categorias de formas. De forma mais específica, investigo se, no processamento da concordância de gênero do português do Brasil, itens regulares quanto ao gênero e as demais formas (transparentes, opacas e irregulares quanto ao gênero) suscitam os mesmos ou diferentes componentes de ERP. Investigo, também, se os ERPs diferenciam-se para processamento da concordância de gênero entre um determinante e um substantivo (condição 1) e entre um substantivo e um adjetivo (condição 2). Para tanto, os sujeitos que participaram da pesquisa foram requisitados a lerem frases congruentes e incongruentes quanto à concordância de gênero nas duas condições. O efeito causado pela diferença dos picos de ERPs gerados para frases congruentes e incongruentes quanto à concordância de gênero nas duas condições (condição 1 e condição 2) foi considerado como uma medida de comparação do processamento da concordância de gênero com substantivos regulares e substantivos das demais formas (cf. MUNTE *et al.*, 1999, NEWMAN *et al.*, 2007).

Na literatura com foco no gênero gramatical, os componentes P600 e LANs foram encontrados em diversos estudos que lidaram com incongruências na concordância de gênero no sintagma nominal tanto na língua materna quanto na segunda língua. Por exemplo, Barber e Carreras (2003) estudaram a diferença entre o processamento do gênero gramatical e o processamento da morfologia de número. Os autores reportaram uma negatividade de 400ms com uma distribuição posterior em respostas a incongruências a violações de gênero e número que está associada a processos de integração lexical.

Foucart e Frenck-Mestre (2011) estudaram a distância estrutural entre os elementos na concordância de gênero. Um P600 foi gerado para incongruências na concordância entre substantivos e adjetivos. Banón, Fiorentino e Gabriele (2012) investigaram em que medida a distância estrutural entre os elementos que concordam em gênero e número afetam o processamento da concordância no Espanhol. As violações na concordância de gênero e número produziram um P600 entre 400ms e 900ms, atingindo sua maior amplitude quando a violação ocorria entre elementos próximos (por exemplo, no sintagma nominal) que entre elementos distantes (violações na concordância de gênero entre sentenças). Outros estudos (GUNTER; FRIEDERICI; SCHRIEFERS, 2000; BARBER; CARREIRAS, 2005; MOLINARO *et al.*, 2008,)

mostram uma LAN para o processamento de incongruências na concordância de gênero.

Na segunda língua, Dowens *et al.* 2011, em um estudo com foco no processamento do gênero do espanhol, também encontraram um P600 para violações de gênero entre falantes do espanhol como segunda língua (falantes nativos do chinês, aprendizes tardios do espanhol). Resultados semelhantes haviam sido encontrados por Morga-Short *et al.* (2010) na investigação dos ERPs gerados como resposta a violações na concordância de gênero de uma língua artificial. A língua artificial foi ensinada aos sujeitos participantes da pesquisa em duas modalidades: explícito (como se aprende em uma sala de aula) e implícito (como em uma imersão no cenário da língua). Os resultados mostraram um N400 para violações entre substantivos e adjetivos em ambos os grupos de aprendizado e um P600 para violações entre determinantes e substantivos em ambos os grupos. Esses resultados sugerem que, na segunda língua, é possível obter padrões de processamento semelhantes aos padrões encontrados na língua materna e, ainda, que estruturas sintáticas diferentes envolvem diferentes mecanismos.

O estudo apresentado neste capítulo foi desenhado com o propósito de detectar os componentes descritos na introdução deste capítulo. Neste estudo, detenho-me na análise dos componentes P600 e LANs por serem esses típicos do processamento morfosintático. Se os ERPs que aparecem como resposta à concordância de gênero com itens regulares diferenciam-se dos ERPs que aparecem como resposta à concordância de gênero com as demais formas (opacas, irregulares e transparentes), então os dados evidenciam que há uma dissociação dos sistemas cognitivos para o processamento e representação do gênero gramatical regular e o processamento do gênero gramatical de outras estruturas fonológicas. Se, por outro lado, os mesmos ERPs aparecem para todas as categorias de forma, mas a amplitude das ondas de ERP para itens regulares distingue-se estatisticamente da amplitude do ERP gerados para outras formas, então os dados evidenciam que a concordância de gênero com todas as categorias de formas é processada pelos mesmos mecanismos neurocognitivos, mas diferenças estatisticamente relevantes em amplitude estão relacionadas à força da representação das propriedades fonológicas dos itens e sua relação com o gênero. Em acréscimo, se a concordância entre determinantes e substantivos e entre substantivos e adjetivos suscitam diferentes componentes de ERPs, então os dados evidenciam que diferenças na estrutura sintática da concordância recrutam diferentes mecanismos de processamento.

É importante salientar que as formas que se distinguem das formas regulares (formas opacas, transparentes e irregulares) serão referenciadas como irregulares neste capítulo. Apesar de estar ciente de que formas irregulares são aquelas cuja terminação contradiz a expectativa de seu gênero (formas masculinas com final *-a*, e femininas com final *-o*), adotei, neste capítulo, essa terminologia pela falta de um melhor termo que abarque as demais categorias de forma em conjunto.

A seguir, descrevo a construção dos estímulos utilizados no experimento eletrofisiológico. Na sequência, descrevo o perfil dos sujeitos e os procedimentos de coleta dos sinais de EEG bem como os instrumentos utilizados na pesquisa. Depois de apresentar toda a metodologia empregada no experimento, descrevo e discuto os resultados obtidos.

6.3 Metodologia

Participantes. Participaram do experimento 16 sujeitos (11 mulheres e 5 homens), com visão normal ou corrigida, entre 21 e 41 anos de idade (média=28,75 anos de idade). Entre os 16 sujeitos, 15 eram destros e um era canhoto. Todos os sujeitos recrutados eram brasileiros falantes nativos do português brasileiro. Todos sujeitos recrutados, com exceção de um, eram estudantes de graduação e pós-graduação (doutorado e pós-doutorado) na *Radboud Universiteit* em Nijmegen, Países Baixos, selecionados pelo programa de intercâmbio *Ciências sem Fronteiras*.

Nenhum dos participantes era bilíngue de infância, mas 15 eram falantes tardios do inglês (entre outras línguas) como segunda língua.

Para a participação no experimento, todos os sujeitos preencheram o questionário *online* do banco de dados de sujeitos do *Max Planck institute for Psycholinguistics*¹⁷ antes da coleta dos dados. Nenhum dos sujeitos reportou desordens neurológicas. Todos os sujeitos recrutados residiam nos Países Baixos há menos de um ano e reportaram usar a língua materna (português do Brasil) diariamente. Os sujeitos receberam €10 (dez euros) pela participação no experimento.

Estímulos. O conjunto de estímulos consistiu em 256 diferentes frases em português brasileiro. Para a construção das frases, duas categorias de forma de substantivos e duas categorias de bigramas (substantivos seguidos de adjetivos) foram selecionados do *Corpus*

¹⁷ www.mpi.nl/ppreg.

Brasileiro disponível no sítio da *Linguateca*. As categorias de forma são:

- a) Substantivos regulares quanto ao gênero (femininos com final *-a*, e masculinos com final *-o*), e irregulares (terminações diferentes de *-a* para o gênero feminino, e diferentes de *-o* para o gênero masculino);
- b) Bigramas regulares (substantivos regulares quanto ao gênero seguidos de adjetivos passíveis de flexão), e bigramas irregulares (substantivos femininos com final diferente de *-a*, e substantivos masculinos com final diferente de *-o* seguidos de adjetivos passíveis de flexão).

Para ambas as categorias de forma de substantivos e de bigramas foram selecionados bigramas femininos e masculinos.

Os substantivos e bigramas foram selecionados do *Corpus Brasileiro* dentro de uma escala de frequência: (cf. Apêndice F). A distribuição da frequência de ocorrência dos substantivos regulares não era estatisticamente diferente da distribuição de frequência de ocorrência dos substantivos irregulares ($t=-1.7067, df=125.746, p\text{-value}=0.09$). Da mesma forma, a frequência de ocorrência dos bigramas regulares não era estatisticamente diferente da frequência de ocorrência dos bigramas irregulares ($t=-0.5815, df=121.515, p\text{-value}=0.562$). A Tabela 16 e Tabela 17, abaixo, exibem a estatística da contagem de frequência dos substantivos e bigramas selecionados.

Tabela 16: Contagem de frequência de ocorrência dos bigramas selecionados (Frequência por milhão).

Frequência dos Bigramas							
	Mín.	Primeiro quartil	Média	Mediana	Terceiro quartil	Máx.	Desvio Padrão
Bigramas regulares	158	281	363.6	331.5	447.8	673	111.07
Bigramas Irregulares	149	265	376.3	378.5	441.8	809	134.93
Teste t	Regulares vs. Irregulares (t=-0.5815, df=121.515, p-value=0.562)						

Tabela 17: Contagem de frequência dos substantivos selecionados (Frequência por milhão).

Frequência dos substantivos selecionados							
	Mín.	Primeiro quartil	Média	Mediana	Terceiro quartil	Máx.	Desvio Padrão
Substantivos regulares	146	233	314	306	393.5	514	100.42
Substantivos Irregulares	139	202	284	264	350.5	507	96.01
Teste <i>t</i>	Regulares vs. Irregulares ($t=-1.7067$, $df=125.746$, $p\text{-value}=0.09$)						

As 256 frases foram construídas a partir dos substantivos e bigramas selecionados. Essas frases foram divididas em oito categorias, cada categoria contendo 32 frases. Essas categorias foram distribuídas em duas condições. As condições de frases eram:

- a) frases congruentes e incongruentes quanto à concordância de gênero entre um determinante e o substantivo regular e entre um determinante e um substantivo irregular (condição 1);
- b) frases congruentes e incongruentes quanto à concordância de gênero entre um substantivo regular e um adjetivo e entre um substantivo irregular e um adjetivo (condição 2).

A Tabela 18, a seguir, apresenta a quantidade de estímulos por categoria e por condição.

Tabela 18: Quantidade de estímulos por categoria e por condição.

Quantidade de estímulos por categoria e por condição.			
Incongruência na concordância de gênero entre determinantes e substantivos (CONDIÇÃO 1)	Categorias	Quantidade de substantivos femininos	Quantidade de substantivos masculinos
	Substantivos Regulares	15	17
	Substantivos Irregulares	18	14
Incongruência na concordância de gênero entre substantivos e adjetivos (CONDIÇÃO 2)	Bigramas Regulares	16	16
	Bigramas Irregulares	16	16
Concordância de gênero congruente (CONDIÇÃO 1)	Categorias	Quantidade de substantivos femininos	Quantidade de substantivos masculinos
	Substantivos Regulares	16	16
	Substantivos Irregulares	16	16
Concordância de gênero congruente (CONDIÇÃO 2)	Bigramas Regulares	16	16
	Bigramas Irregulares	16	16

Exemplos de frases incongruentes na concordância de gênero na condição 1:

- a) com substantivos regulares: *As fotos revelaram *o natureza do caso*
- b) com substantivos irregulares: *Os atletas controlaram *os fontes de energia*

Exemplos de frases incongruentes na concordância de gênero na condição 2:

- a) com substantivos regulares: *O documentário mostrou a* obra clássica do pintor*
- b) com substantivos irregulares: *O vapor representa a *fase gasoso da água.*

Exemplos de frases congruentes quanto à concordância de gênero na condição 1:

- a) com substantivos regulares: *O paciente interrompeu o tratamento de saúde*
- b) com substantivos irregulares: *Os trabalhadores receberam a metade das férias*

Exemplos de frases congruentes quanto à concordância de gênero na condição 2:

- a) com substantivos regulares: *A planilha informou o lucro bruto do mês.*
- b) com substantivos irregulares: *O técnico mudou o esquema tático do time.*

Todas as frases da condição 1 e condição 2 apresentaram a mesma quantidade de elementos: condição 1 com sete vocábulos, e condição 2 oito vocábulos. Todas as frases foram construídas usando a mesma estrutura sintática:

Condição 1: Determinante (artigo definido ou indefinido) – substantivo – verbo – objeto direto (substantivo) – complemento nominal;

Condição 2: Determinante (artigo definido ou indefinido) – substantivo – verbo – objeto direto (substantivo seguido de adjetivo) – complemento nominal.

As frases construídas a partir dos substantivos e bigramas selecionados foram submetidas a um teste de naturalidade em duas etapas. Na primeira etapa, o conjunto de estímulos foi submetido a cinco falantes nativos do português que indicaram quais frases lhes pareciam estranhas ou naturais. Após esse julgamento, as frases julgadas como estranhas foram alteradas. Na segunda etapa, o conjunto de estímulos (incluindo as frases alteradas) foi então submetido a outros dez falantes nativos do português do Brasil que julgaram a naturalidade das frases com base em uma escala de sete pontos, sendo 1 ponto ‘muito estranha’ e 7 ‘muito natural’. Após o julgamento, a média da pontuação foi calculada para cada uma das frases. Dado que, nessa segunda etapa, nenhuma das frases apresentou uma média abaixo de cinco pontos, todas foram mantidas no conjunto de estímulos apresentados aos sujeitos (Apêndice F).

O experimento foi programado usando a linguagem de programação *Presentation*TM.

Procedimentos. Depois de assinarem o termo de consentimento livre e esclarecido de participação no experimento (cf. Apêndice G), todos os sujeitos preencheram o questionário da base de dados de sujeitos do *Max Planck Institute for Psycholinguistics in Nijmegen*. Em seguida, uma toca elástica foi aderida ao escalpo dos sujeitos para a colocação dos eletrodos. Na sequência, os sujeitos foram conduzidos a uma cabine à prova de som para dar início à sessão experimental.

A sessão experimental foi iniciada com um treino. Esse treino contou com a apresentação de dez frases congruentes na concordância de gênero as quais não faziam parte do conjunto de estímulos do experimento. Após o treino, os sujeitos foram instruídos a lerem atentamente as frases que apareciam na tela do computador e a não piscarem durante a leitura das frases.

O experimento foi programado ser apresentado em quatro blocos de 64 frases: oito frases de cada condição ($8 \times 8 = 64$). Após a apresentação de cada frase, uma pausa de dois segundos foi introduzida entre frases para que os sujeitos piscassem. Um intervalo para descanso foi introduzido entre a apresentação dos blocos. Os sujeitos foram instruídos a descansarem pelo tempo que achassem necessário.

A ordem de apresentação dos estímulos de cada bloco bem como a ordem de apresentação dos blocos foram randomizadas entre os sujeitos para evitar efeitos de ordem.

Os vocábulos das frases foram apresentados visualmente, um de cada vez no centro da tela. O tempo de apresentação de cada vocábulo foi de 300ms e levou em consideração o comprimento ortográfico do estímulo: 37ms por caracter foi acrescido ao tempo de apresentação de cada vocábulo (cf. LAI; CURRAN, 2013). O intervalo entre palavras foi de 100ms. A apresentação de cada frase era iniciada com um sinal de fixação que permanecia na tela por 500ms. A palavra alvo da condição 1 (substantivo) era sempre o quinto vocábulo apresentado, e a palavra alvo da condição 2 (adjetivo) era sempre o sexto vocábulo apresentado.

Para o controle da atenção dos sujeitos na leitura, seis perguntas do tipo SIM/NÃO¹⁸ foram apresentadas em cada bloco, 24 perguntas no total (Apêndice F). As perguntas versavam sobre o conteúdo de frases congruentes quanto à concordância de gênero. Essas perguntas foram apresentadas imediatamente após o aparecimento da frase congruente.

Para resposta SIM, os sujeitos foram instruídos a pressionarem o botão verde da caixa de resposta, e para resposta NÃO os sujeitos foram instruídos a pressionarem o botão vermelho da caixa de respostas. Os sujeitos não receberam *feedback* de seu desempenho nas respostas. Em média, cada sessão experimental durou 2 horas, incluindo o tempo de preparação dos sujeitos.

Vale ressaltar que nenhum sujeito foi informado do objetivo do experimento anteriormente à coleta de dados. Ao final da coleta, cada um dos sujeitos recebeu um *debriefing* (Apêndice H) que apresentava os objetivos do experimento e a relação desse objetivo com o desenho experimental adotado.

6.4 Procedimento de coleta de dados de ERP e análise dos dados

A atividade elétrica cerebral dos sujeitos foi coletada a partir de 59 eletrodos ativos equidistantes montados em uma toca elástica (EasyCap manufaturada para a montagem de 64 canais) e amplificada usando o *BrainAmp DC*.

A saída do amplificador foi digitalizada com uma taxa de amostragem (*sampling rate*) de 500Hz. Um eletrodo (eletrodo 61) foi posicionado abaixo do olho esquerdo para o controle de movimentos verticais (piscadas), e os eletrodos posicionados na lateral externa de cada olho (eletrodos 25 e 57) serviram de controle dos movimentos horizontais dos olhos. Um eletrodo (eletrodo 32) foi posicionado no

¹⁸ Essas perguntas não foram incluídas nas análises.

mastoide direito, a referência implícita no mastoide esquerdo. Os dados foram pré-processados e analisados usando o *Brain Vision Analyzer* (versão 2.0). Todos os eletrodos foram re-referenciados ao mastoide direito *off-line* antes da extração das médias de ERP. A impedância foi mantida abaixo de 5Ω (cinco *ohms*). O protocolo¹⁹ do experimentador consta no Apêndice I desta tese.

A remoção de artefatos musculares e movimentos dos olhos foi realizada em modo semi-automático. Primeiramente, um algoritmo baseado em regressão (ICA - *Independent Component Analysis algorithm*) detectou os picos de piscadas e movimentos horizontais. Em seguida, o desempenho do algoritmo na marcação dos artefatos foi verificado por meio de uma inspeção visual. Artefatos não detectados pelo algoritmo foram removidos manualmente, e os artefatos detectados inapropriadamente foram mantidos.

Antes da extração das médias globais dos ERPs de cada categoria de estímulo, todos os canais de cada sujeito foram filtrados. O filtro Butterworth passa baixas frequências de 30 Hz e o filtro passa-altas frequências de 0,05Hz foram aplicados aos sinais. Após a aplicação dos filtros, os sinais foram divididos em épocas de 1100 ms, incluindo uma *baseline* de -200ms pré-estímulo.

As médias das amplitudes dos ERPs coletados para substantivos e bigramas regulares e irregulares das frases na condição 1 e condição 2 congruentes e incongruentes foram calculadas para duas janelas de latência: de 300ms a 500ms para detectar as negatividades (LANs), e de 500ms a 900ms para analisar o efeito das positividades tardias (P600).

A estatística foi realizada em linguagem *R*, método ANOVA (2 x 2 x 2 x 3 x 2) medidas repetidas, pacote *ez* (LAWRENCE, 2013). Para a estatística das negatividades e positividades, os fatores incluídos foram: *Forma* (regular vs. irregular), *Congruência* (Congruentes vs. Incongruentes) *Condição* (Condição 1 vs. Condição 2), *Posição* (Anterior vs. Posterior vs. Central), *Hemisfério* (Direito vs. Esquerdo).

A Tabela 19, a seguir, apresenta os eletrodos incluídos na análise estatística de cinco regiões do escalpo: hemisfério direito anterior e posterior; hemisfério esquerdo anterior e posterior; eletrodos da linha central.

A Figura 14 ilustra a montagem dos eletrodos e as cinco regiões do escalpo incluídas na estatística.

¹⁹ Disponível o protocolo original escrito em inglês.

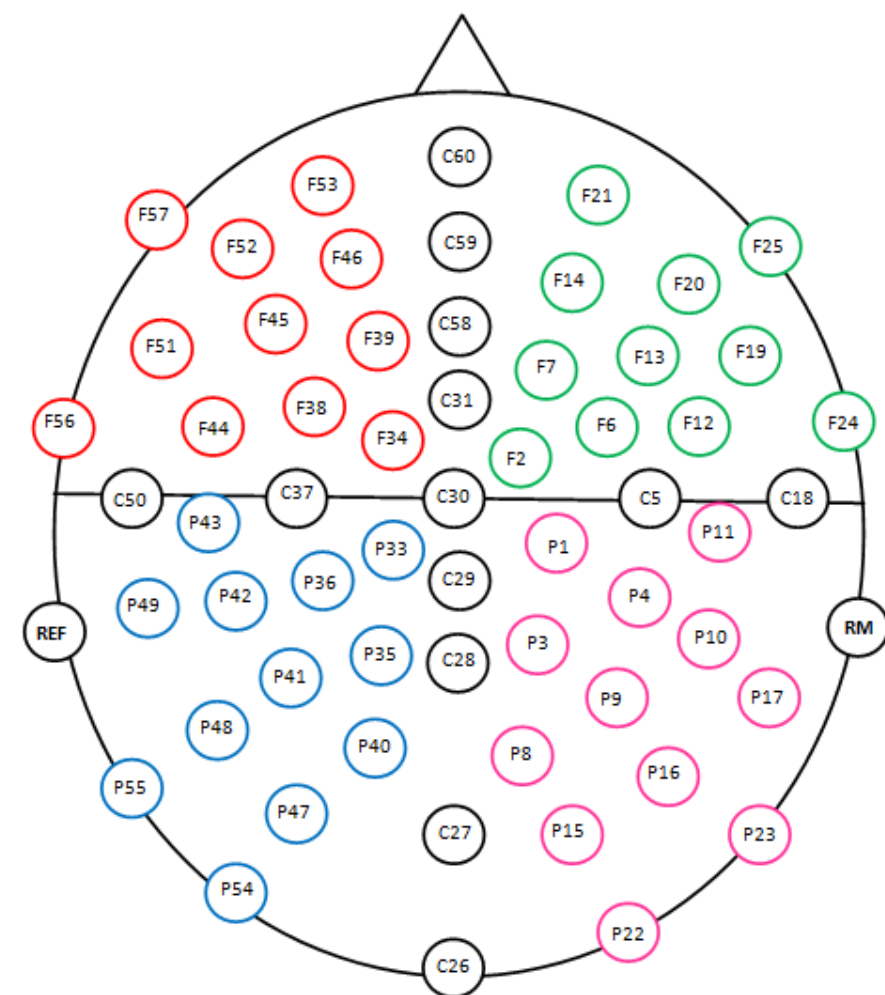
O teste ANOVA medidas repetidas foi conduzido em todas as cinco regiões do escalpo, bem como em cada uma das cinco regiões separadamente em ambos os intervalos de tempo: entre 300ms e 500ms e entre 500ms e 900ms. O teste Geisser e Greenhouse de esfericidade foi aplicado nas medidas repetidas com mais de um grau de liberdade. Os resultados reportados apresentam os graus de liberdade originais e os valores corrigidos de p .

Tabela 19: Eletrodos incluídos na estatística para cinco regiões do escalpo

Hemisfério	Posição	Eletrodos
Esquerdo	Anterior	F34, F39, F46, F53, F38, F44, F45, F52, F57, F51, F56
Esquerdo	Posterior	P40, P47, P54, P35, P41, P48, P55, P33, P36, P42, P49, P43
Direito	Anterior	F2, F6, F7, F14, F21, F13, F20, F25, F19, F12, F24
Direito	Posterior	P1, P4, P10, P17, P11, P3, P9, P16, P23, P22, P15, P8
Central	Central	C31, C58, C59, C60, C26, C27, C28, C29, C30, C5, C18, C50, C37

Hemisfério esquerdo anterior _____
 Hemisfério direito anterior _____
 Hemisfério esquerdo posterior _____
 Hemisfério direito posterior _____
 Linha Central _____

Figura 14: Montagem dos eletrodos incluídos na estatística.
 F: eletrodos frontais; P: eletrodos posteriores; C: eletrodos centrais.



6.5 Resultados

A Figura 15, Figura 16 e Figura 17 exibem as médias globais dos ERPs extraídos para frases congruentes e incongruentes na concordância de gênero com substantivos regulares e irregulares (condição 1), e a Figura 18, Figura 19 e Figura 20 exibem as médias globais dos ERPs extraídos para frases congruentes e incongruentes na concordância de gênero com bigramas regulares e irregulares (condição 2). Figura 21 e Figura 22 apresentam a distribuição topográfica dos componentes de ERP gerados para concordância de gênero com substantivos regulares e irregulares na condição 1 (Figura 21) e condição 2 (Figura 22) em épocas anteriores (entre 300ms e 500ms). Figura 23 e Figura 24 apresentam a distribuição topográfica dos componentes de ERP gerados para concordância de gênero com substantivos regulares e irregulares na condição 1 (Figura 23) e condição 2 (Figura 24) em épocas posteriores (entre 500ms e 900ms).

O mapa topográfico das voltagens representa a diferença entre os picos de ERPs gerados para sentenças congruentes e incongruentes de cada categoria de forma (regulares vs. irregulares) em cada condição (condição 1 e condição 2).

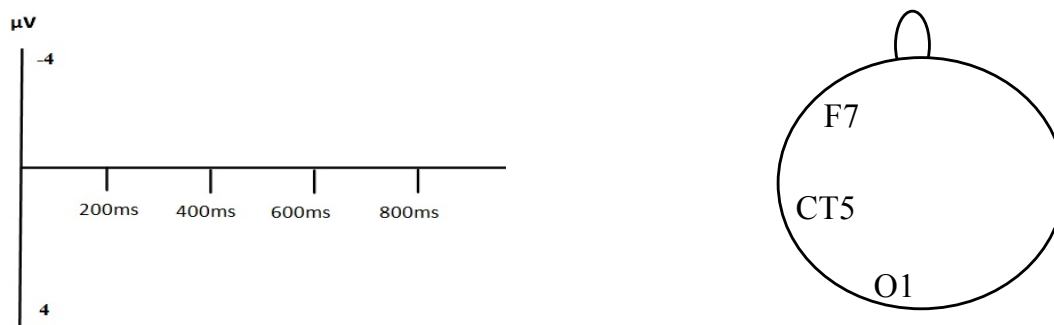
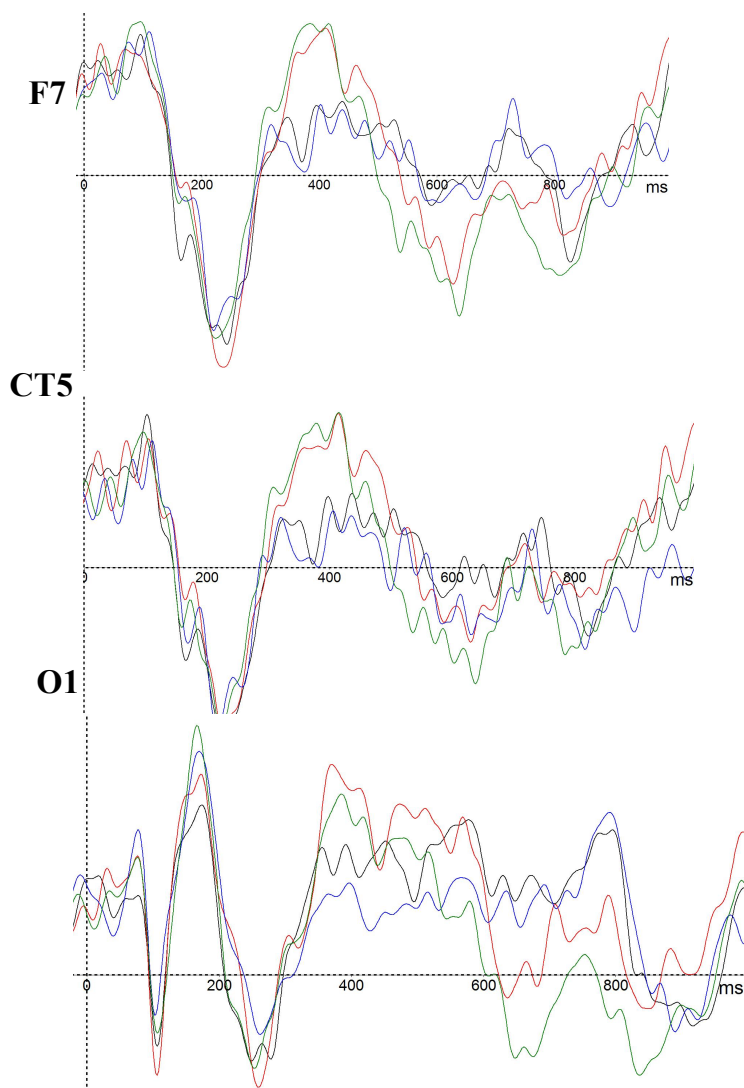


Figura 15: Médias globais dos ERPs gerados para a concordância de gênero entre determinantes e substantivos (condição 1). Hemisfério esquerdo. Ondas negativas para cima e positivas para baixo.



Incongruente irregular _____ Incongruente regular _____
 Congruente regular _____ Congruente irregular _____

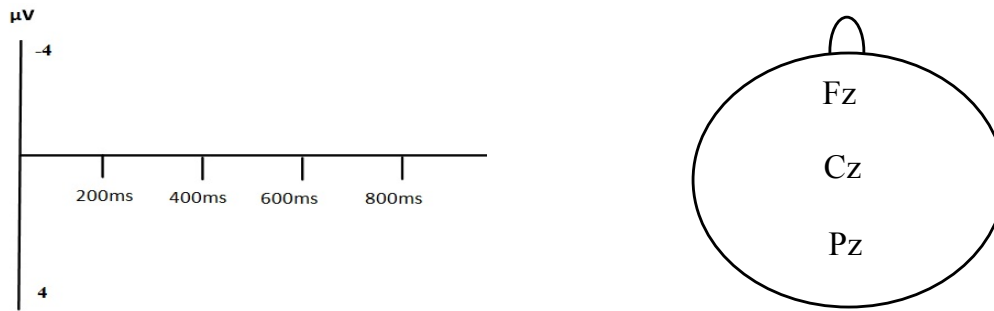
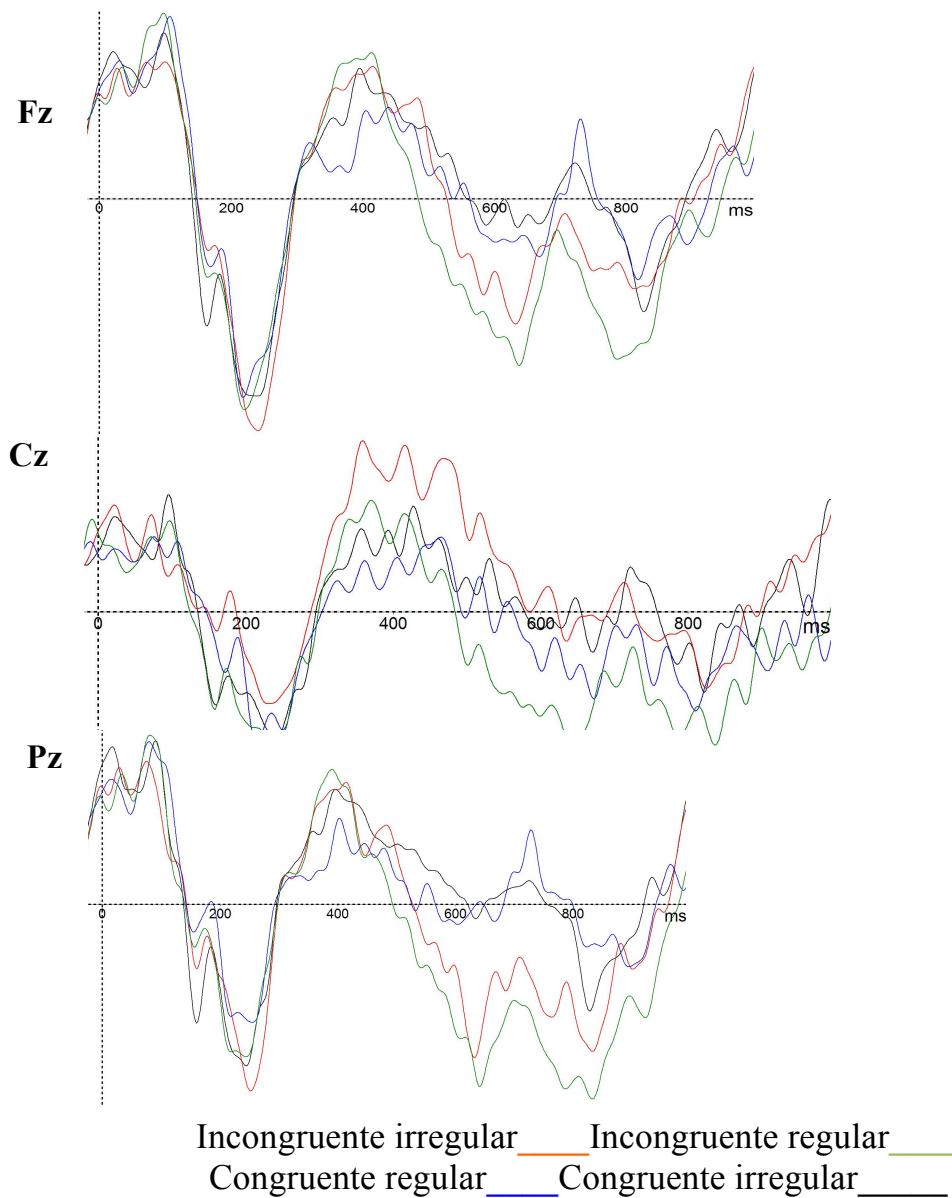


Figura 16: Médias globais dos ERPs gerados para a concordância de gênero entre determinantes e substantivos (condição 1). Linha central. Ondas negativas para cima e positivas para baixo.



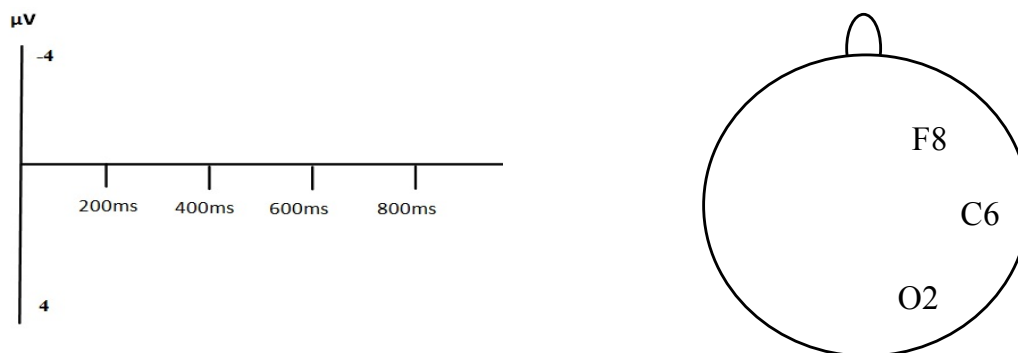
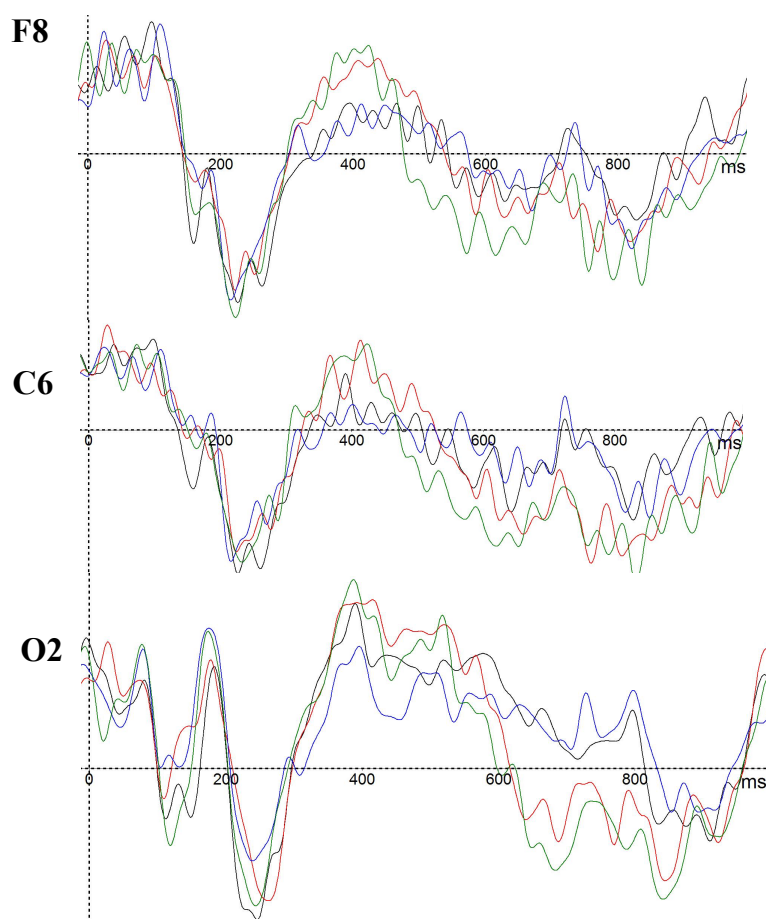


Figura 17: Médias globais dos ERPs gerados para a concordância de gênero entre determinantes e substantivos (condição 1). Hemisfério direito. Ondas negativas para cima e positivas para baixo.



Incongruente irregular _____ Incongruente regular _____
 Congruente regular _____ Congruente irregular _____

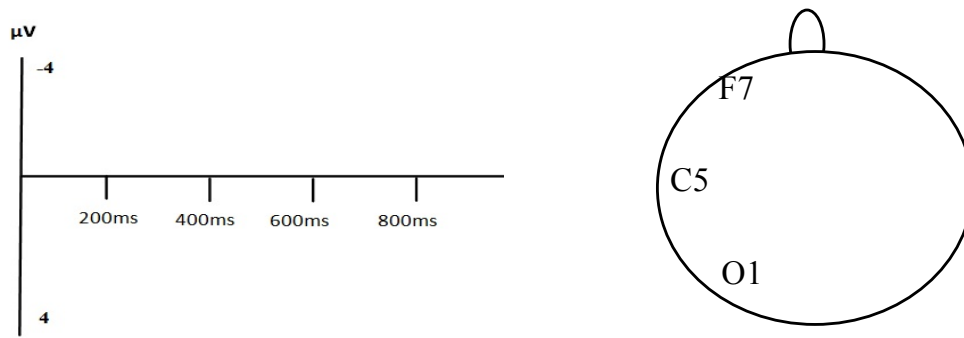
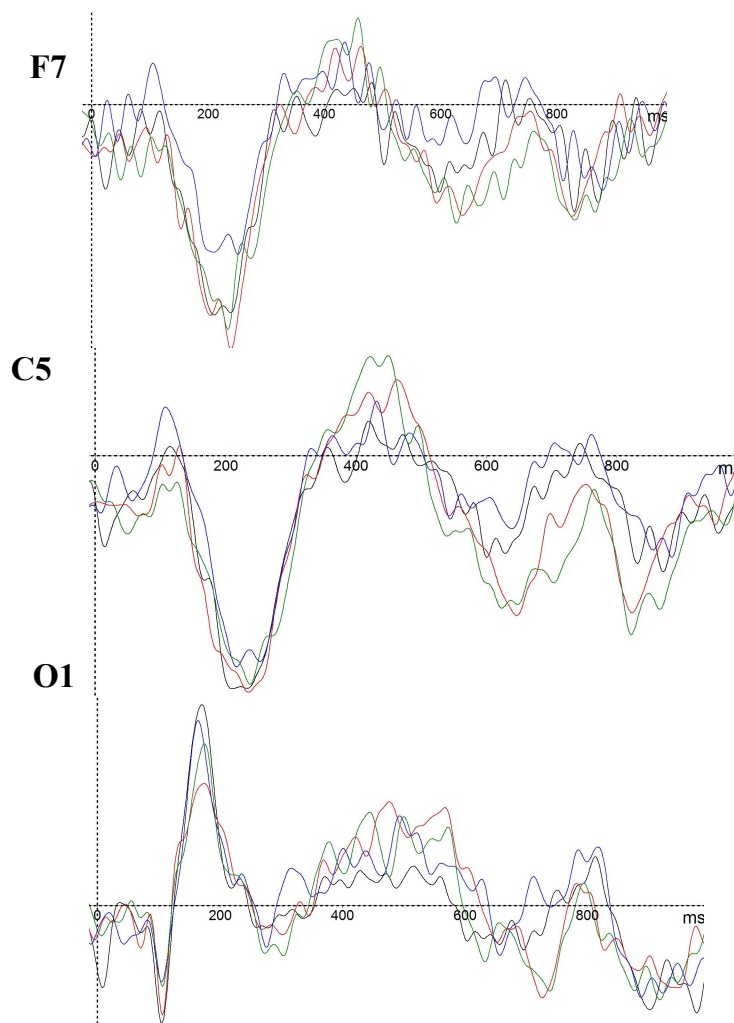


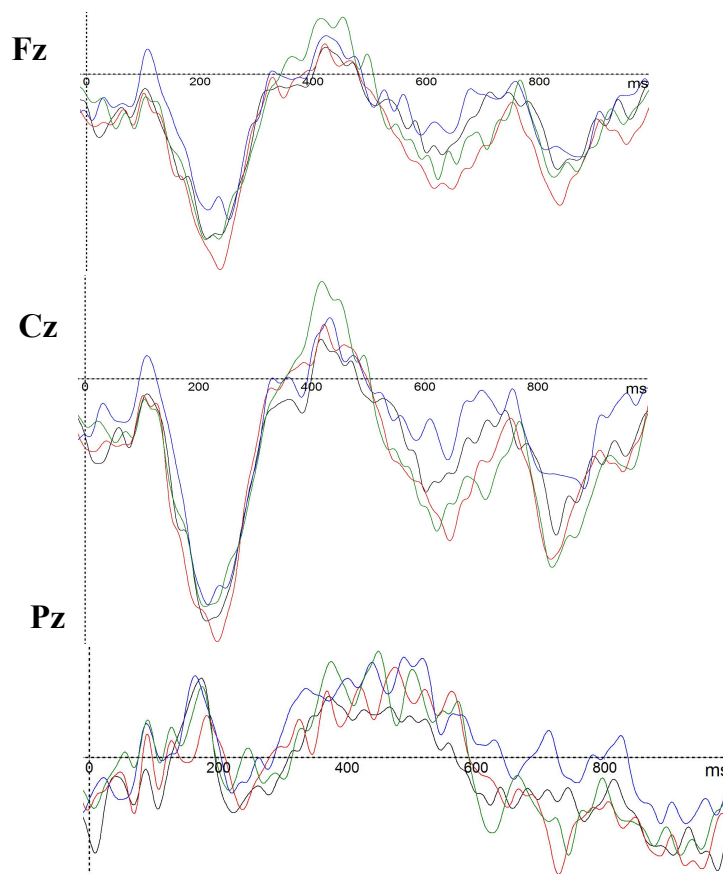
Figura 18: Médias globais dos ERPs gerados para a concordância de gênero entre substantivos e adjetivos (condição 2). Hemisfério esquerdo. Ondas negativas para cima e positivas para baixo.



Incongruente irregular _____ Incongruente regular _____
 Congruente regular _____ Congruente irregular _____



Figura 19: Médias globais dos ERPs gerados para a concordância de gênero entre substantivos e adjetivos (condição 2). Linha central. Ondas negativas para cima e positivas para baixo.



Incongruente irregular _____ Incongruente regular _____
 Congruente regular _____ Congruente irregular _____

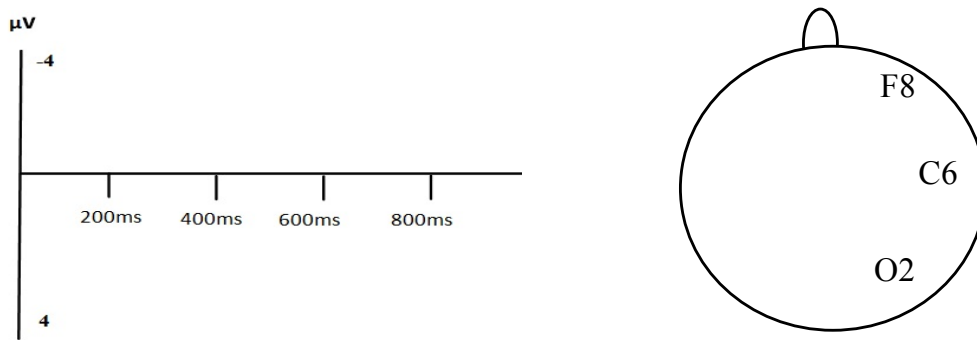
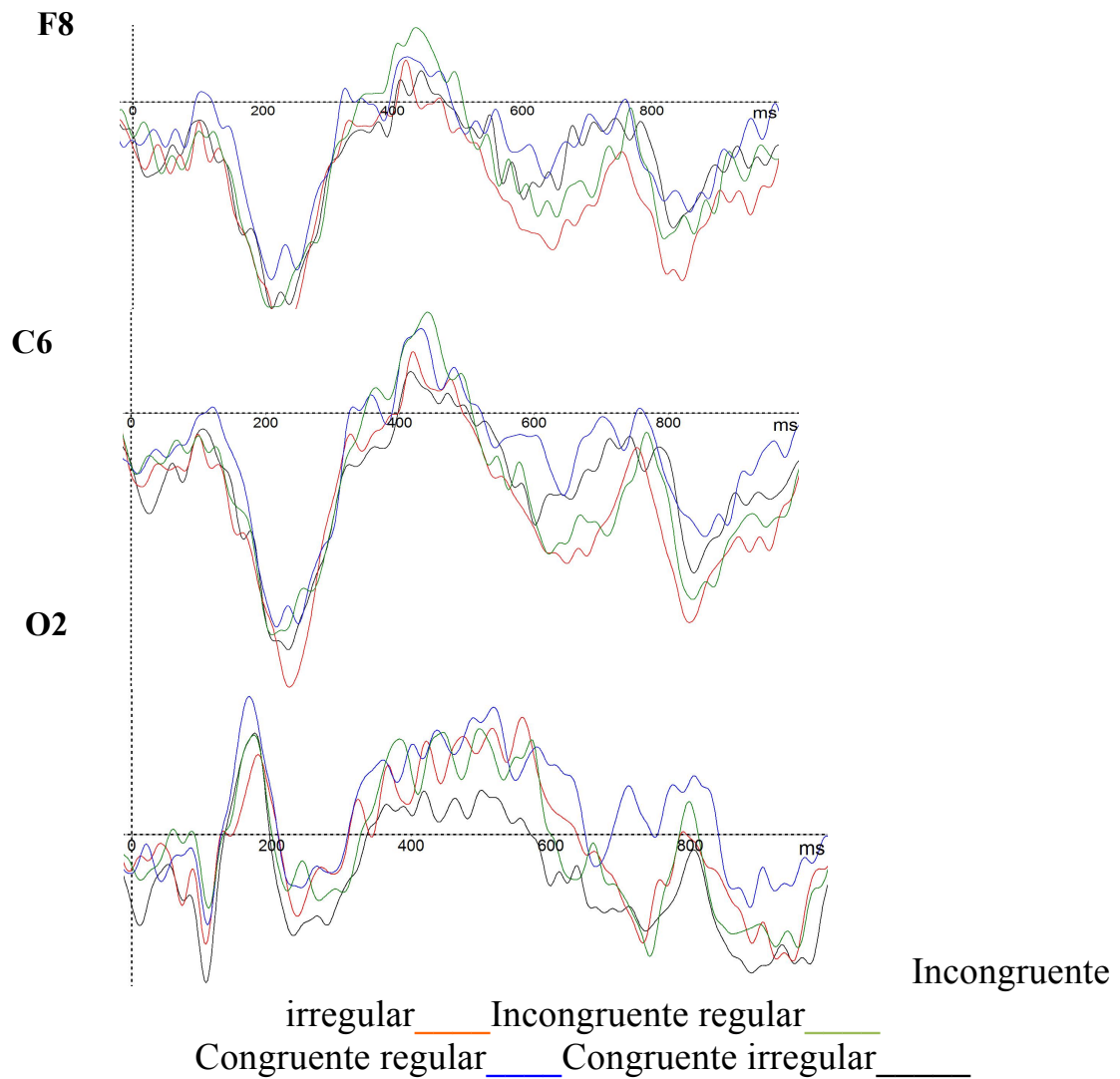


Figura 20: Médias globais dos ERPs gerados para a concordância de gênero entre substantivos e adjetivos (condição 2). Hemisfério direito. Ondas negativas para cima e positivas para baixo.



Negatividades – Condição 1	
Substantivos regulares	Substantivos Irregulares

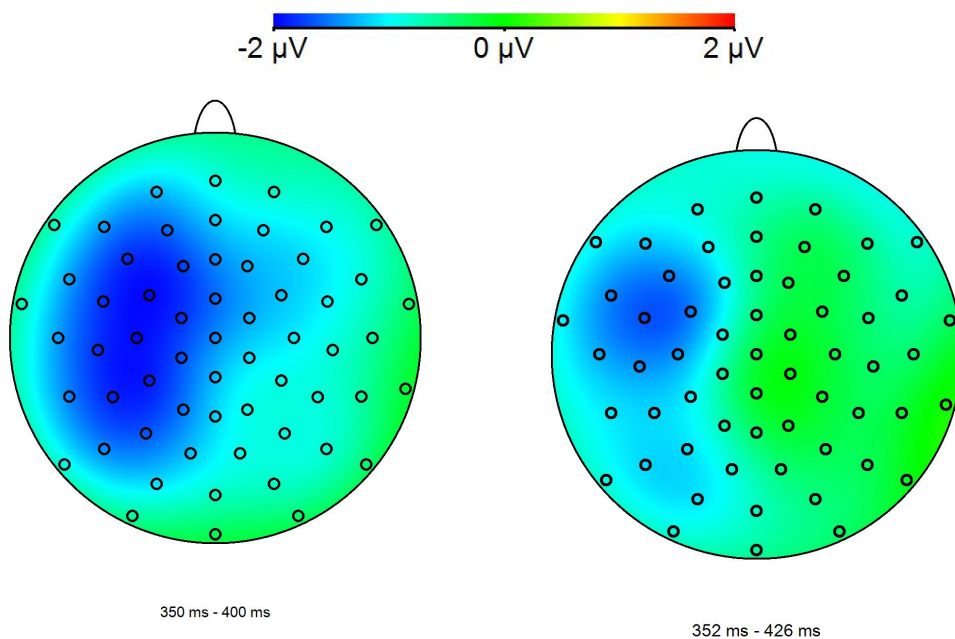


Figura 22: Mapas topográficos do escalpo mostrando a distribuição dos ERPs nas épocas anteriores (entre 300ms e 500ms). Foram escolhidas janelas de 50ms para capturar a diferença de pico de ERP entre sentenças congruentes e incongruentes em cada categoria de forma na condição 2.

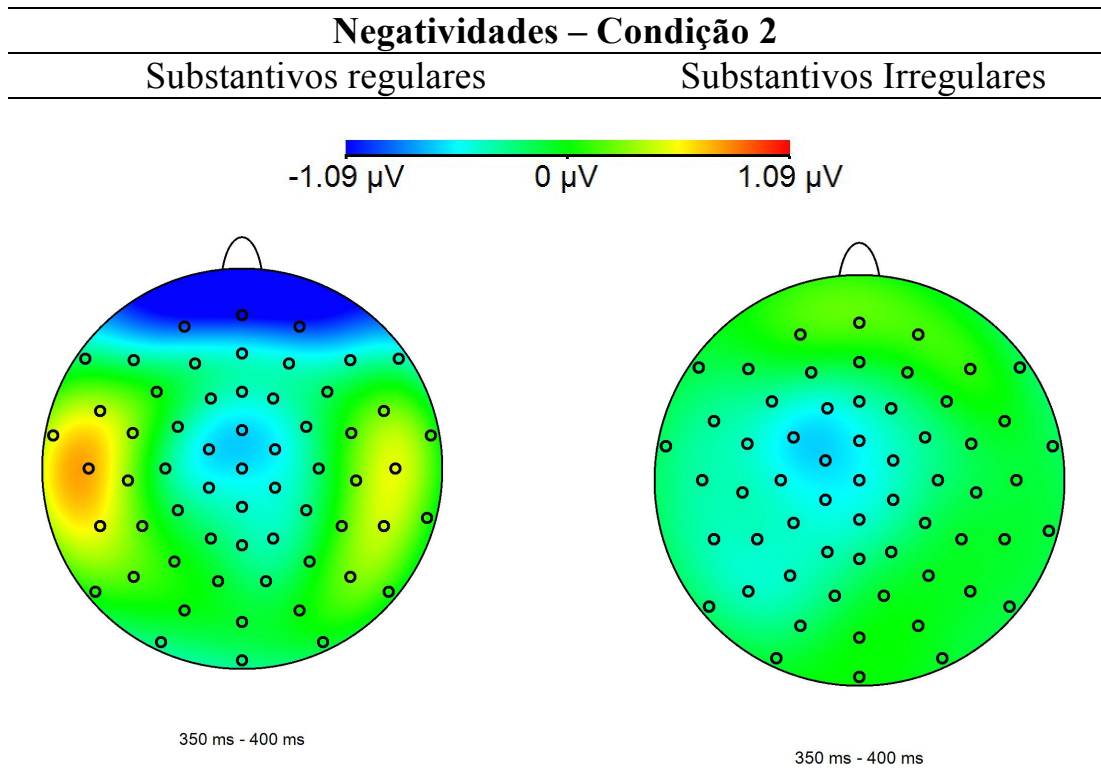


Figura 23: Mapas topográficos do escalpo mostrando a distribuição dos ERPs nas épocas posteriores (entre 500ms e 900ms). Foram escolhidas janelas de 50ms para capturar a diferença de pico de ERP entre sentenças congruentes e incongruentes em cada categoria de forma na condição 1.

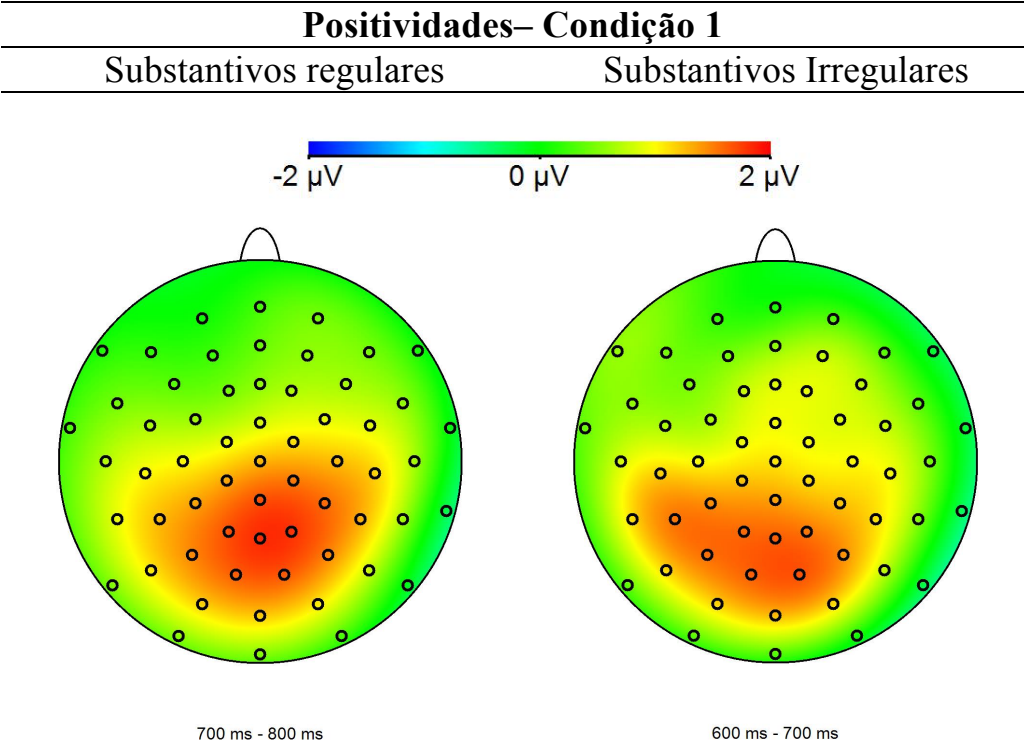
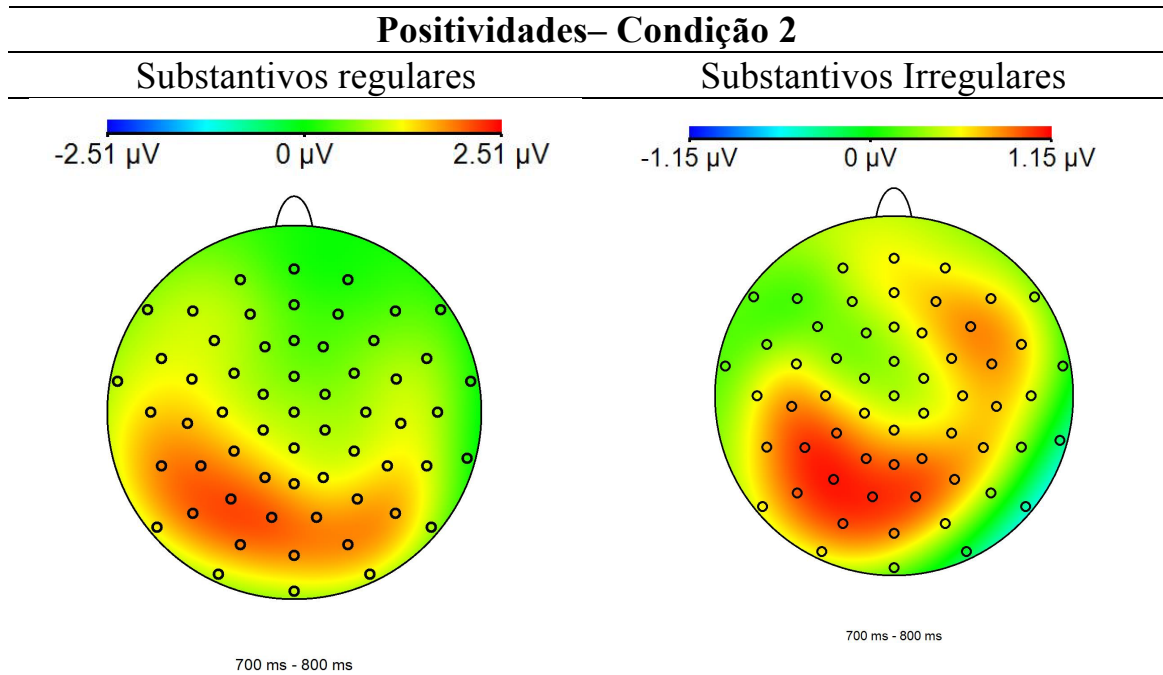


Figura 24: Mapas topográficos do escalpo mostrando a distribuição dos ERPs nas épocas posteriores (entre 500ms e 900ms). Foram escolhidas janelas de 50ms para capturar a diferença de pico de ERP entre sentenças congruentes e incongruentes em cada categoria de forma na condição 2.



300ms a 500ms.

A inspeção visual da Figura 15 revela uma negatividade entre 300ms e 500ms que atinge seu pico máximo no hemisfério esquerdo na condição 1 (eletrodos F7 e CT5) para incongruências na concordância de gênero. Na condição 2, Figura 19 mostra uma negatividade entre 300ms e 500ms na linha central do escalpo (eletrodo Cz) para incongruências na concordância de gênero. A negatividade na condição 2 é mais enfatizada na concordância de gênero com substantivos regulares. Note que incongruências na concordância de gênero com substantivos regulares e irregulares, tanto na condição 1 quanto na condição 2, apresentam uma amplitude bastante aproximada. Essa proximidade foi confirmada pelo teste ANOVA que não revelou efeito significativo do fator forma (regular x irregular) moldando as amplitudes, nem uma interação entre os fatores forma, hemisfério e incongruência. Essa ausência do efeito do fator forma e ausência de interação com outros fatores indica que o processamento de substantivos regulares e irregulares não é diferente estatisticamente em nenhuma região do escalpo no intervalo de tempo entre 300 e 500ms, nem se diferenciam quando associado ao fator congruência e hemisfério.

A distribuição topográfica das negatividades observadas na condição 1 e condição 2 revela que, na condição 1, tanto substantivos regulares quanto irregulares suscitaram uma LAN (Figura 21), sugerindo que ambas as categorias de formas acionaram mecanismos cognitivos próprios de processos automáticos de concordância morfossintática. No entanto, na condição 2, o efeito LAN não foi encontrado (Figura 22).

Para investigar o efeito LAN da condição 1 estatisticamente, um teste ANOVA foi conduzido para a região anterior do hemisfério esquerdo. O teste ANOVA conduzido para essa região revelou um efeito independente significativo do fator *Congruência* ($F(1,15)=4,75$, $p=0,04$). Testes *t* pareados conduzidos no hemisfério esquerdo mostraram que os ERPs gerados por frases incongruentes são estatisticamente diferentes dos ERPs gerados por frases congruentes apenas na condição 1, tanto para substantivos regulares ($t = -4.4201$, $df = 346.628$, $p<0,01$) quanto para substantivos irregulares ($t = -4.5447$, $df = 334.489$, $p<0,01$). Na condição 2, por sua vez, o teste *t* pareado conduzido para a análise do hemisfério esquerdo revelou que os ERPs gerados por frases incongruentes não foram estatisticamente diferentes dos ERPs gerados por frases congruentes nem para a concordância com substantivos regulares ($t = -1.0931$, $df = 349.532$, $p\text{-value} = 0.3$), nem para a concordância com substantivos irregulares ($t = -1.8608$, $df = 349.992$, $p\text{-value} = 0.06$).

500ms a 900ms.

A inspeção visual das Figura 15-20 revela positivities entre 600ms e 700ms e entre 800ms e 900ms tanto na condição 1 quanto na condição 2. É possível observar, também, que, tanto na condição 1 quanto na condição 2, há diferenças de amplitude entre a concordância de gênero congruente e incongruente.

Na condição 1, a positividade gerada pela concordância de gênero incongruente com substantivos regulares apresenta maior amplitude que a positividade gerada para a concordância incongruente com substantivos irregulares (Figuras 15-17). O pico máximo da positividade gerada para ambas as categoria de forma de substantivo em frases incongruentes da condição 1 aparece nos eletrodos da regiões centro-parietais (Figura 16 - eletrodos Pz e Cz). Ainda na condição 1, a concordância congruente com substantivos regulares apresenta ondas positivas com amplitude aproximada à amplitude da onda da concordância congruente com substantivos irregulares em todas as regiões do escalpo.

Na condição 2, a inspeção visual das Figura 18-20, revela que a amplitude das positivities geradas para frases incongruentes com substantivos regulares aproxima-se da amplitude da positividade gerada por frases incongruentes com substantivos irregulares em todas as regiões do escalpo. Por outro lado, diferentemente da condição 1, a diferença entre as amplitudes das positivities geradas por frases congruentes e incongruentes com substantivos regulares é maior que a diferença entre as amplitudes da positividade gerada por frases congruentes e incongruentes com substantivos irregulares, principalmente na região posterior do hemisfério esquerdo (eletrodos Pz e O1 e O2), indicando que o efeito do fator congruência foi maior para a concordância regular que para a concordância irregular.

Os mapas topográficos de voltagens dos componentes de ERPs revelam que as positivities apresentam uma distribuição no escalpo compatível com a distribuição do P600 tanto na condição 1 quanto na condição 2, concentrando principalmente nas regiões centrais e posteriores do escalpo. O efeito P600 (diferença entre os picos de positivities geradas para frases congruentes e incongruentes) foi encontrado na condição 1 tanto para o processamento da concordância de gênero com substantivos regulares quanto com substantivos irregulares, indicando que ambas as categorias de forma acionaram os mesmos mecanismos neurocognitivos associados a processos de re-análise e reparo da estrutura (Figura 23).

Na condição 2, o efeito P600 aparece como relevante apenas para a concordância com substantivos regulares (Figura 24), indicando diferenças na ativação de mecanismos entre as condições (condição 1 e condição 2), bem como entre as categorias de forma (regular vs. Irregular) na condição 2. Estatisticamente, as diferenças entre as condições foi revelado pelo teste ANOVA conduzido para as regiões posteriores do escalpo, especialmente para a região posterior do hemisfério esquerdo: o teste ANOVA revelou um efeito do fator condição ($F(1,15)=4,59$, $p=0,04$), indicando diferenças no processamento da concordância de gênero entre a condição 1 e condição 2.

As diferenças observadas entre as amplitudes da positividade geradas para a concordância congruente e incongruente na condição 1 e condição 2 bem como diferenças entre as condições no processamento da concordância de gênero regular e irregular (condição 1 e condição 2) foi estatisticamente revelado pelo teste ANOVA conduzido na porção posterior do hemisfério esquerdo. O teste ANOVA revelou um efeito do fator condição ($F(1,15)=4,59$, $p=0,04$), um efeito do fator congruência ($F(1,15)=7,4$, $p=0,01$) e um efeito da interação dos fatores congruência e forma ($F(1,15)=7,5$, $p=0,01$). Esses efeitos indicam, respectivamente, que a condição 1 diferencia-se da condição 2, sentenças congruentes diferenciam-se estatisticamente de incongruentes em ambas as condições e que, como sugere a inspeção visual das Figura 18-20, na condição 2, a diferença entre as amplitudes das positividade geradas por frases congruentes e incongruentes é maior para a concordância com substantivos regulares que para a concordância com substantivos irregulares.

Cabe ressaltar que a interação entre os fatores forma e incongruência foi encontrada apenas na porção posterior do hemisfério esquerdo. Essa diferença da região posterior do hemisfério esquerdo em relação às outras regiões foi revelado pelo teste ANOVA conduzido para todas as regiões em conjunto. Nesse teste, o fator posição foi estatisticamente relevante ($F(2,30)=9,59$, $p=0,002$).

6.6 Discussão dos resultados

Neste capítulo, um experimento eletrofisiológico foi conduzido com o objetivo de investigar se o processamento da concordância de gênero com substantivos regulares e irregulares suscita os mesmos ou diferentes sinais eletrofisiológicos. O objetivo secundário do

experimento foi investigar se duas condições de concordância de gênero suscitam os mesmos ou diferentes componentes de ERP.

Para cumprir os objetivos propostos, os sujeitos participantes do experimento foram apresentados a frases congruentes e incongruentes quanto à concordância de gênero entre determinantes e substantivos regulares e irregulares (condição 1), e frases congruentes e incongruentes na concordância de gênero entre substantivos regulares e irregulares e adjetivos (condição 2).

A diferença entre os picos de ERPs gerados por frases congruentes e incongruentes entre 300ms e 500ms revelaram um efeito bifásico LAN/P600 na condição 1 tanto para a concordância de gênero regular quanto para a concordância de gênero irregular. Esses resultados indicam que o processamento da concordância de gênero entre determinantes e substantivos regulares e irregulares suscitaram os mesmos sinais eletrofisiológicos. Conforme menciona Friederici (2002), diferenças na distribuição espacial, assim como diferenças na polaridade da atividade elétrica sugerem o envolvimento de diferentes estruturas neurais e, por isso, diferenças no processamento. Portanto, é possível concluir que ambas as categorias de forma acionaram os mesmos mecanismos neurocognitivos. Em acréscimo, o efeito bifásico LAN/P600 indica que o processamento da concordância de gênero com ambas as categorias de forma na condição 1 envolve o recrutamento de processos automáticos de processamento morfossintático (cf. MOLINARO; BARBER; CARRERAS, 2011).

Uma crítica pode ser levantada em relação a essa interpretação: o efeito LAN da condição 1 para a concordância regular apresentou uma duração maior que o efeito LAN para o processamento da concordância irregular bem como apresentou maior magnitude de distribuição no hemisfério esquerdo em oposição a uma LAN suscitada pela concordância irregular com distribuição mais focal na região anterior do hemisfério esquerdo (cf. Figura 21). No entanto, cumpre esclarecer que a interpretação aqui apresentada está baseada no fato de o aparecimento do efeito LAN, tanto para regulares quanto para irregulares, ter sido refletido estatisticamente por meio de um efeito independente do fator congruência (frases congruentes vs. incongruentes), indicando que tanto o processamento da concordância regular e irregular foi influenciado pelo fator congruência na condição 1.

Diferentemente da condição 1, nenhum componente foi encontrado entre 300ms e 500ms para nenhuma categoria de forma na condição 2. A ausência do componente LAN ou qualquer outro componente foi comprovado pelo teste *t* o qual revelou que, na porção

anterior do hemisfério esquerdo, diferenças estatísticas entre as concordâncias congruentes e incongruentes ficaram restritas à condição 1. A ausência do componente LAN na condição 2 para ambas as categorias de forma de substantivo sugere que, na concordância de gênero entre substantivos e adjetivos, processos automáticos de análise morfossintática não estão envolvidos.

Na condição 2, o efeito P600 encontrado foi significativo apenas para a concordância com formas regulares. Esses resultados sugerem que, quando diferentes formas de substantivos estão envolvidas, diferentes sinais eletrofisiológicos podem emergir. Contudo, o efeito P600 encontrado para a concordância com substantivos regulares indica que o processamento da concordância de gênero entre substantivos e adjetivos recruta mecanismos de re-análise, já que o P600 está associado a dificuldades de integração de uma estrutura no contexto anterior (cf. KAAAN, 2007).

Quais as implicações desses resultados para os arcabouços teóricos discutidos nesta tese, a saber: os modelos teóricos de via dual e modelos teóricos de via única?

Em primeiro lugar, o efeito bifásico LAN/P600 suscitado tanto pela concordância com substantivos regulares quanto pela concordância com substantivos das demais formas (tratadas aqui neste capítulo como irregulares) na condição 1 não fornece evidências para as teorias de processamento dual. No experimento aqui apresentado, nem mesmo uma diferença entre a amplitude da concordância regular e irregular foi estatisticamente relevante em nenhum dos testes ANOVA conduzidos para o intervalo de tempo entre 300ms e 500ms em nenhuma região do escalpo analisada. Por outro lado, o efeito P600 encontrado apenas para formas regulares na condição 2 foi manifestado estatisticamente tanto pelo efeito do fator condição, sugerindo diferenças estatísticas entre as condições de concordância, quanto pelo efeito da interação entre os fatores congruência e forma encontrado para a região posterior do hemisfério esquerdo, indicando que diferenças de amplitude de ERPs gerados por sentenças congruentes e incongruentes ficaram restritas a positivities suscitadas por formas regulares quanto ao gênero apenas na região posterior do hemisfério esquerdo na condição 2. Assim, na condição 2, mecanismos de re-análise e reparo foram recrutados apenas para a concordância com formas regulares. Entretanto, em minha interpretação, esse dado não fornece evidências suficientes para concluir que a concordância entre substantivos regulares e adjetivos é processada por mecanismos diferentes. Essa interpretação está fundamentada no fato de o P600 ser, como já mencionado, um componente de ERP

associado a processos não automáticos. Conforme apontado por Friederici, Hahne e Saddy (2002), o P600 é reflexo de processos não automáticos de revisão e reparo da estrutura sintática.

É possível que o efeito P600 tenha sido gerado apenas para a concordância regular em virtude de incongruências na concordância de gênero com substantivos regulares causarem um impacto maior nos sujeitos que incongruências na concordância com substantivos transparentes, opacos e irregulares, ativando com mais facilidade um processo de re-análise e reparo que incongruências com as demais formas. É possível que o impacto causado pelas incongruências com substantivos regulares esteja relacionado ao fato da terminação *-a* ser fortemente associada ao gênero feminino, e a terminação *-o* fortemente associada ao gênero masculino em comparação às outras terminações do português. Nesse sentido, parece-me que o efeito P600 da condição 2 está mais associado com a força da representação de *-a* para o gênero feminino e *-o* para o gênero masculino que ativação de diferentes mecanismos.

Em resumo, tendo em vista os resultados aqui apresentados, concluo que os dados apontam em direção a um processamento do gênero em uma única rota. Vale ressaltar que os resultados obtidos estão em consonância com os resultados reportados na literatura para o processamento do gênero gramatical para outras línguas que também encontraram um efeito bifásico LAN/P600 para incongruências na concordância de gênero, principalmente no que se refere à concordância de gênero entre determinantes e substantivos (cf. Seção 6.2).

PARTE III

PARTE III: RESUMO E CONCLUSÕES

PARTE III

PARTE III

Neste último capítulo, apresento um resumo dos capítulos anteriores ressaltando os objetivos específicos de cada um. Em seguida, descrevo a conclusão geral desta tese, tendo em vista os resultados obtidos em todos os experimentos em conjunto. Por fim, apresento as limitações do trabalho e forneço ideias para pesquisa futuras com foco gênero gramatical do português. As sugestões de pesquisas futuras estão baseadas nas limitações do trabalho.

CAPÍTULO 7

7 RESUMO E CONCLUSÕES

7.1 Resumo

O objetivo geral desta tese foi buscar um melhor entendimento dos processos mentais que envolvem o gênero gramatical da variedade brasileira do português. O objetivo específico desta tese foi investigar se o processamento do gênero do português é governado por uma única ou por uma dupla rota. Para cumprir os objetivos, uma série de experimentos foi conduzida a fim de buscar dados para que questões relativas ao processamento e representação mental do gênero do português pudessem ser elucidadas.

No Capítulo 2 desta tese, apresentei uma descrição da relação existente entre gênero gramatical e a estrutura fonológica dos nomes substantivos de referente animado e inanimado assim como características da concordância de gênero do português. A partir da relação observada entre forma e gênero dos substantivos inanimados do português, foi possível distribuí-los em quatro categorias de forma: regulares (nomes femininos terminados em *-a*, e masculinos terminados em *-o*), irregulares (nomes femininos terminados em *-o*, e masculinos terminados em *-a*), transparentes (terminações típicas do gênero feminino como *-agem* e *-ade*, e típicas do gênero masculino *-ume*, *-ude*, *or*) e opacas quanto ao gênero (*noite*, *tarde*, *lápiz*). A distribuição dos substantivos em categorias de forma foi importante para que o processamento de cada categoria de forma pudesse ser investigado levando em conta as premissas dos modelos teóricos de processamento da linguagem de interesse para esta tese as quais foram apresentadas no Capítulo 3 desta tese.

No Capítulo 3, vimos que os modelos de via dual (ULLMAN; 1999, 2001; PINKER; 1991, 1999) assumem que formas regulares, passíveis de decomposição, são processadas na gramática mental e formas irregulares processadas no léxico mental. Segundo a visão dual de processamento, a gramática mental representa um mecanismo neurocognitivo especializado na computação de regras. O léxico mental, por sua vez, seria um sistema de memória responsável pelo armazenamento de itens idiossincráticos em sua forma e pelo estabelecimento de associações entre formas que compartilham traços semânticos, fonológicos, ortográficos e morfológicos.

Os modelos de via única e modelos conexionista, por outro lado, apresentam uma perspectiva divergente. Na perspectiva de

processamento unitário, todas as formas são processadas no léxico mental, isto é, em um único sistema de memória de natureza associativa. Os modelos de via única e, em particular, o *Network Model*, assumem, ainda, que propriedades estruturais mais frequentes apresentam maior força associativa que propriedades estruturais menos frequentes. As asserções teóricas desses dois modelos balizaram o desenho experimental adotado nos experimentos apresentados nos Capítulo 4, Capítulo 5 e Capítulo 6.

No Capítulo 4, os resultados de três experimentos comportamentais visualmente apresentados foram reportados. No Experimento 1, o processamento da concordância de gênero entre determinantes e formas regulares e opacas foi contrastado. No Experimento 2, outras categorias de forma (formas transparentes e irregulares) foram incluídas em uma tarefa de atribuição de gênero em que os sujeitos deveriam decidir se um substantivo inanimado era feminino ou masculino. No Experimento 3, pseudopalavras com terminações típicas de formas regulares, transparentes, opacas e irregulares quanto ao gênero do português foram apresentadas aos sujeitos para que esses executassem uma tarefa de atribuição de gênero.

No Experimento 1, o Modelo de Regressão Linear de Efeitos Mistos revelou um efeito da interação dos fatores *Frequência* e *Forma*, indicando que formas regulares e opacas foram influenciadas pela frequência dos itens, mas de maneira distinta. Em uma análise exploratória dos dados, testes *t* pareados revelaram um efeito maior da frequência para formas regulares que para formas opacas quanto ao gênero. Os resultados obtidos no Experimento 1 sugerem, por um lado, que a decisão de gênero para ambas as categorias de forma são baseadas na memória de seu gênero e, por outro lado, a ativação do sistema computacional para os itens regulares infrequentes.

Os resultados do Experimento 2 revelaram um efeito individual da frequência influenciando todas as formas. Revelaram, ainda, um aumento gradativo na velocidade de resposta dos sujeitos: Os sujeitos foram mais rápidos em média, em primeiro lugar, para decidir o gênero de formas regulares, em segundo lugar, para decidir o gênero de formas transparentes, em terceiro lugar, para decidir o gênero de formas opacas e, por último, de formas irregulares. Esses resultados sugerem novamente que o fator frequência cumpre papel fundamental na decisão do gênero dos substantivos. Sugerem, ainda, que, mais que forma, o que está em jogo é a força da representação de uma propriedade estrutural e sua associação com uma categoria de gênero.

Os resultados do Experimento 3 com pseudopalavras mostraram que, independentemente da terminação da pseudopalavra, quanto mais traços estruturais uma forma compartilha com outras pré-existentes no léxico mental dos sujeitos, mais fácil é o acesso lexical, consequentemente, mais rápida é a velocidade de resposta dos sujeitos na decisão do gênero. Os resultados do Experimento 3 sugerem, portanto, o acionamento de um mecanismo associativo para todas as categorias de forma.

Evidências para o acionamento desse mecanismo associativo foram encontradas, também, na comparação das porcentagens de classificação de pseudopalavras ‘fáceis’ e ‘difíceis’ em categorias de gênero: as porcentagens de acurácia na classificação das pseudopalavras fáceis em categorias de gênero foi, em geral, superior que as porcentagens de acurácia da classificação das pseudopalavras ‘difíceis’ em categorias de gênero. Além disso, o tempo de reação a pseudopalavras fáceis foi, de modo geral, inferior quando comparado ao tempo de reação a pseudopalavras difíceis. De acordo com o *Network model*, itens lexicais interconectados a outros itens por meio de propriedades estruturais compartilhadas são mais facilmente acessados e, por isso, apresentam menor tempo de reação.

No capítulo 5, três simulações computacionais foram conduzidas usando um *perceptron* multicamadas do tipo *feed forward* com treinamento utilizando o algoritmo *backpropagation*. O banco de dados para o treinamento incluiu formas regulares e irregulares quanto ao gênero, formas opacas femininas e masculinas e formas transparentes quanto ao gênero. O modelo que apresentou melhor desempenho numa validação cruzada foi escolhido para as três simulações empreendidas. A rede neural foi treinada com 75% do banco de dados e os restantes 25% do conjunto de dados foi separado para a Simulação 1. O modelo resultante do treinamento do algoritmo com 75% dos dados apresentou uma acurácia estimada em 100%. Na Simulação 1, o modelo conexionista foi capaz de classificar os 25% dos dados em categorias de gênero com uma acurácia de 100%. Na simulação 2, o modelo selecionado foi usado para prever o gênero de seis pseudopalavras que compartilhavam traços estruturais com cinco formas opacas (terminação em *-e*) femininas e masculinas presentes no treinamento da rede. O algoritmo apresentou uma acurácia de 80%, apenas um erro de classificação, indicando que, independentemente da ambiguidade da terminação, o algoritmo apresentou um desempenho satisfatório e baseou-se nos traços compartilhados com estruturas incluídas no treinamento.

Resultados semelhantes foram encontrados na Simulação 3. Nessa simulação, o desempenho do modelo foi testado com seis pseudopalavras que compartilhavam traços estruturais com formas ambíguas quanto ao gênero (femininas e masculinas com terminação em *-a*). Novamente, o modelo apresentou um desempenho satisfatório com uma acurácia de 83% (apenas um erro de classificação de gênero). Os resultados tanto da Simulação 2 quanto da Simulação 3 indicam que os erros na classificação da rede estão relacionados com a força de representação de uma estrutura e sua associação com uma categoria de gênero. Nesse sentido, um sistema conexionista de via única ‘imitou’ o comportamento dos sujeitos participantes do Experimento 3 do Capítulo 4, já que tanto a rede neural quanto os sujeitos levaram em consideração a estrutura das pseudopalavras como um todo e não apenas a terminação para a decisão do gênero. Juntos, os resultados das simulações sugerem que, assim como os experimentos conduzidos com humanos, as decisões de gênero de um modelo conexionista são baseadas nos traços compartilhados entre as estruturas, um processo tipicamente associativo.

No capítulo 6, foram apresentados os resultados de um experimento eletrofisiológico que teve como principal objetivo investigar (e contrastar) os sinais de ERP suscitados pelo processamento da concordância de gênero com substantivos regulares e pelo processamento da concordância de gênero com as demais formas (referenciadas no capítulo como irregulares). Para tanto, 16 sujeitos, falantes nativos do português do Brasil, foram solicitados a lerem frases do português congruentes e incongruentes na concordância de gênero entre um determinante e um substantivo (condição 1), e entre um substantivo e um adjetivo (condição 2). O efeito produzido pela diferença entre os picos de ERP gerados pelo processamento de frases congruentes e incongruentes na concordância de gênero com ambas as categorias de formas (regulares e irregulares) nas duas condições (condição 1 e condição 2) foi tomado como medida de comparação do processamento da concordância de gênero regular e irregular. Os resultados revelaram um efeito bifásico LAN/P600 na condição condição 1 para ambas as categorias de forma, mas apenas um efeito P600 para o processamento de formas regulares na condição 2.

Os sinais eletrofisiológicos suscitados bem como dos testes estatísticos conduzidos mostraram que: i) o processamento da concordância de gênero regular e irregular suscita os mesmos componentes de ERP, sugerindo o recrutamento dos mesmos mecanismos neurocognitivos; ii) que não há diferenças estatisticamente significativas entre as amplitudes dos ERPs suscitadas para formas

regulares e irregulares; iii) ao contrário do processamento entre um determinante e um substantivo, o processamento da concordância de gênero entre um substantivo e um adjetivo regular ou irregular não recruta mecanismos associados ao processamento automático da concordância de gênero, já que o P600 é uma positividade tardia associada a processos de re-análise e reparo da estrutura (morfo)sintática.

7.2 Conclusões

Os resultados dos estudos empíricos apresentados nos Capítulos 4 a 6 mostraram que diferenças entre o processamento do gênero dos substantivos regulares, irregulares, transparentes e opacos não podem ser explicadas com base apenas nas diferenças entre as formas dos substantivos inanimados do português. Os dados dos experimentos comportamentais e das simulações computacionais mostraram que, mais que forma, a memória cumpre papel fundamental no processamento do gênero gramatical do português independentemente da terminação do item lexical. Esses achados estão em consonância com os resultados de alguns estudos que trataram do impacto dos fatores forma e frequência no processamento do gênero gramatical, principalmente, estudos com focalizando no processamento do gênero em francês (TAFT; MEUNIER, 1998; DESROCHERS *et al.*, 1989). Em acréscimo, os dados provenientes do experimento eletrofisiológico mostraram que tanto formas regulares quanto as demais estruturas fonológicas recrutam LANs, um componente típico de processos automáticos de concordância morfossintática. As simulações computacionais mostraram que um simples sistema conexional de via única é capaz de estabelecer associações entre as formas para atribuir-lhes gênero. A Simulação 1 mostrou que o modelo lida com formas regulares e irregulares com alto grau de acurácia. As Simulações 2 e 3 mostraram que o modelo é capaz de estabelecer associações entre uma palavra e uma pseudopalavra com base nos traços compartilhados entre as estruturas. Os resultados das Simulações 2 e 3 indicam também que o modelo conexional leva em conta a frequência de ocorrência de uma propriedade estrutural e sua relação com uma categoria de gênero. Portanto, um modelo conexional de via única foi capaz de reproduzir aspectos do processamento do gênero do português observados nos resultados dos experimentos comportamentais apresentados no Capítulo 4 desta tese.

De posse dos resultados, concluo que a hipótese inicial de que o processamento do gênero gramatical é governado por uma dupla rota

não pôde ser confirmada. A única pista para a ativação de um mecanismo computacional foi encontrada para o processamento de pseudopalavras e processamento de formas muito infrequentes. Assim, os dados aqui obtidos apresentam evidências que permitem concluir que os pressupostos dos modelos de via dual são simplistas para explicar o processamento do gênero do português, pois não dão conta de explicar a ativação de um mecanismo associativo para formas regulares, nem mesmo a ativação de um mecanismo computacional somente para palavras muito infrequentes (como visto no Experimento 1 do Capítulo 4) e pseudopalavras (Experimento 3 do Capítulo 4). Portanto, concluo que os dados das amostras em exame podem ser melhor explicados pela visão connexionista e unitária de processamento, em especial, pela visão de processamento morfológico segundo o *Network Model*. Assim sendo, com base nas evidências empíricas aqui apresentadas, sustento que o processamento do gênero do português é melhor representado por um único sistema (neuro)cognitivo. Mais especificamente, sustento que todas as formas, inclusive as regulares quanto ao gênero do português, são armazenadas no léxico mental e, por isso, todas são passíveis de entrarem em um processo associativo. Nesse processo associativo, propriedades estruturais mais frequentes apresentam maior força associativa que propriedades estruturais menos frequentes.

Cabe ressaltar, no entanto, que as conclusões aqui apresentadas estão limitadas à análise de dados provenientes de experimentos conduzidos visualmente. Estudos eletrofisiológicos e comportamentais usando tarefas com estímulos auditivos, bem como um modelo connexionista simulando a apresentação auditiva de substantivos para classificação em categorias de gênero seriam interessantes para verificar se os resultados obtidos nesta tese podem ser generalizados para os resultados obtidos por experimentos auditivos. Outros estudos interessantes com foco no gênero do português seriam:

- c) Estudos contrastando o processamento do gênero natural (gênero de substantivos de referente animado) com o processamento do gênero gramatical para investigar se o processamento do gênero natural distingue-se do processamento do gênero gramatical em virtude de sua associação com a informação de sexo do referente;
- d) Assim como fizeram Taft e Meunier (1998) para o francês, estudos com o objetivo de investigar em que medida a informação contida no determinante favorece o processamento do gênero gramatical de substantivos inanimados;

- e) Estudos comparando os sinais eletrofisiológicos suscitados por formas regulares, irregulares, opacas e transparentes quanto ao gênero;
- f) Estudos combinando fMRI e EEG para investigar a resolução temporal e espacial do processamento do gênero de substantivos inanimados do português;
- g) Uma simulação computacional conexionista do processamento do gênero com uma maior quantidade de dados, incluindo substantivos de referente animado. Esse modelo teria que ser treinado não só com informações da estrutura e sua relação com uma categoria de gênero gramatical, mas, também, com informações de traços semânticos, como: referente humano/não humano; animado/não animado.

Referências

AFONSO, Olivia; DOMÍNGUEZ, Alberto; ÁLVAREZ, Carlos J.; MORALES, David. Sublexical and Lexico-Syntactic Factors in Gender Access in Spanish. *Journal of Psycholinguistic research*, 2013.

ALEGRE, Maria; GORDON, Peter. Frequency Effects and the representational status of regular inflections. *Journal of Memory and Language*, v. 40, pp. 41-61, 1999.

ANDERSON, Stephen R. Where's morphology. *Linguistic Inquiry*, v. 13, n. 4, pp.571-612, 1982.

AZEREDO, José Carlos de. *Gramática Houaiss da Língua Portuguesa*. São Paulo, Publifolha, 2013.

BANÓN, José Alemán; FIORENTINO, Robert; GRABIELE, Alison. The processing of number and gender agreement in Spanish: An event-related potential investigation of the effects of structural distance. *Brain Research*, v. 1456, pp. 49-63, 2012.

BAAYENS, Harald; DIJKSTRA, Ton; SCHREUDER, Robert. Singulars and Plurals in Dutch: Evidence for a parallel dual route model. *Journal of Memory and Language*, v. 37, pp. 94-117, 1997.

BAAYEN, Harald R., TWEEDIE, Fiona J., SCHREUDER, Robert. The subjects as a simple random effect fallacy: Subject variability and morphological family effects in the mental lexicon, *Brain and Language*, 81, pp.55–65, 2002.

BAAYEN, Harald R.; SCHREUDER, Robert; DE JONG, Nivja; KROTT, Andrea. Dutch inflections: The rules that prove the exceptions. In. NOOTEBOOM, Sieb.; WEERMAN, Fred.; WIJNEN, Frank.(Orgs). *Storage and Computation in the Language Faculty*, Dordrecht: Kluwer Academic Publishers, pp. 61–92, 2002.

BATES, Elisabeth; MACWHINNEY, Brian. Competition, variation, and language learning. In: MACWHINNEY, Brian (Org.), *Mechanisms of language acquisition*. Hillsdale, NJ: Lawrence Erlbaum Associates, 1989.

- BATES, Elisabeth. A., WULFECK, Brian. Crosslinguistic studies of aphasia. In: MACWHINNEY, Brian; BATES, Elisabeth (Orgs.). *The crosslinguistic study of sentence processing*. New York: Cambridge University Press, pp. 328-371, 1989.
- BATES, Elisabeth. A., WULFECK, Brian. Crosslinguistic studies of aphasia. *Brain and Language*, 41, p. 123-148, 1991.
- BATES, Elisabeth; DEVESCOVI, Antonella; PIZZAMIGLIO, Luigi; D'AMICO, Simona; HERNADEZ, Arturo. Gender and lexical access in Italian. *Perception & Psychophysics*, v.57, pp. 847—862, 1995.
- BATES Douglas, MAECHLER, Martin, Bolker, Ben M.; Walker Steven. “lme4: Linear mixed-effects models using Eigen and S4.” *Journal of Statistical Software*, 2014.
Disponível em: <URL: <http://arxiv.org/abs/1406.5823>>.
- BARBER, Haracio; CARRERAS, Manuel. Integrating Gender and Number information Spanish word pairs: An ERP study, *Cortex*, v. 39, pp. 465-482, 2003.
- BARBER, Horacio; SALILLAS, Elena; CARREIRAS, Manuel. Gender or genders agreement? In: CARREIRAS, Manuel; CLIFTON, C. (Orgs.): *Libro: On-line study of sentence comprehension; Eyetracking, ERP and beyond*. Psychology Press, pp. 309-328, 2004.
- BARBER, Horacio; CARREIRAS, Manuel. Gramatical gender and number agreement in Spanish: An ERP comparison. *Journal of Cognitive Neuroscience*, v.17, n.1, pp. 137-153, 2005.
- BARR, Dale J.; LEVY, Roger; SCHEEPERS, Christoph; TILLY, Harry J.. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, v. 68, pp. 255–278, 2013.
- BECHARA, Evanildo. *Moderna Gramática Portuguesa*. Nova Fronteira, 37ª edição, 2009.
- BIEN, Heidrun. On the production of morphologically complex words with special attention to effects of frequency. 145 f. Tese (Doutorado) - Radbound University, Nijmegen, 2007.

BOWDEN, Harriet Wood. *Proficiency and Second-Language Neurocognition: A Study Of Spanish as a first and second language*. 241 f. Tese (Doutorado) - Georgetown University, Washington D.C., 2007.

BROVETTO, Claudia; ULLMAN, Michael. T. The Mental Representation and Processing of Spanish Verbal Morphology. In: EDDINGTON, David (Org.), *Selected Proceedings of the 7th Hispanic Linguistics Symposium*, Somerville, Cascadilla Press, pp.98-105, 2005.

BYBEE, Joan L. *Morphology: a study of the relation between meaning and form*. Philadelphia: John Benjamins, 1985.

BYBEE, Joan L. Regular Morphology and the Lexicon. *Language and Cognitive Processes*, v. 10, pp. 425-455, 1995.

CÂMARA JR., Joaquim Mattoso. *Estrutura da língua portuguesa*. Petrópolis: Vozes, 1970.

CHATER, Nick., MANNING, Christopher. Probabilistic models of language processing and acquisition. *Trends in Cognitive Science*, v.10, n.7, pp. 335-344, 2006.

CHOMSKY, Noam. *The Minimalist Program*. Cambridge, MA.: MIT Press, 1995.

CHOMSKY, Noam. *Lectures on government and binding*. Dordrecht: Foris, 1981.

CLAHSEN, Harald; ROTHWEILER, Monika; WOEST, Andreas. Regular and irregular in the acquisition of German noun plurals. *Cognition*, v. 45, pp. 225-255, 1992.

MARCUS, Gary F.; BRINKMANN, Ursula; CLAHSEN, Harald; WIESE, Richard; PINKER, Steven. German inflection: The exception that proves the rule. *Cognitive Psychology*, n. 29, pp. 189-256, 1995.

CLAHSEN, Harald; EISENBEISS, Sonja; SONNENSTUHL, Ingrid. Morphological structure and the processing of inflected words, *Theoretical Linguistics*, v. 23, pp. 201-249, 1997.

CORBETT, Greville. *Gender*. Nova York: Cambridge University Press, 1991.

CORRÊA, Letícia Maria Sicuro. Concordância de gênero no processamento de formas pronominais. *Cadernos de Estudos Linguísticos*. v. 40, pp.77-92, Jan/Jun., 2001.

COULSON, Seana; KING, Jonathan W.; KUTAS, Martha. Expect the Unexpected: Even- Related Brain Response to Morphosyntactic Violations. *Language and Cognitive Processes*, v. 13, n.1, pp. 21-58, 1998.

CUNHA, Celso; CINTRA, Lidney. *Nova Gramática do Português Contemporâneo*. Rio de Janeiro: Nova Fronteira, 2001.

DESROCHERS, Alain; PAIVIO, Allan; DESROCHERS, Sylvie. L'effet de la fréquence d'usage des noms inanimés et de la valeur prédictive de leur terminaison sur l'identification du genre grammatical. *Revue Canadienne de Psychologie*, v. 43, pp.62-73, 1989.

DOWENS, Margaret Gillon; GUO, Taomi; GUO, Jingjing; BARBER, Horacio; CARREIRAS, Manuel. Gender and number processing in Chinese Learners of Spanish – Evidence from Event Related Potentials. *Neuropsychologia*, v.49, pp.1651-1659,2011.

FLEISCHHAUER, Elisabeth; CLAHSEN, Harald. Generating Inflected Word Forms in Real Time: Evaluating the Role of Age, Frequency, and Working Memory. Proceedings of the 36th annual Boston University Conference on Language Development. Somerville: Cascadilla Press, 2012.

FOSTER, Kenneth; CHAMBERS, Susan M. Lexical access and naming time. *Journal of verbal learning and verbal behavior*.v. 12, pp. 627-635, 1973.

FOUCART, Alice; FRENCK-MESTRE, Cheryl. Can late L2 learners acquire new grammatical features? Evidence from ERPs and eye-tracking. *Journal of Memory and Language*, v. 66, pp.226-248, 2012.

FRIEDERICI, Angela D.; PFEIFER, Erdmut; HAHNE, Anja.. Event-related brain potentials during natural speech processing: effects of semantic, morphological and syntactic violations. *Cognitive Brain Response*, v.1, pp. 183–192, 1993.

FRIEDERICI, Angela D.; HAHNE, Anja; SADDY, Douglas. Distinct Neurophysiological Patterns Reflecting Aspects of Syntactic Complexity and Syntactic Repair. *Journal of Psycholinguistic Research*, v.31, n.1, January 2002.

GOLLAN, Tamar. H.; FROST, Ram. Two routes to grammatical gender: Evidence from Hebrew. *Journal of Psycholinguistic Research*, v. 30, p. 627–651, 2001.

GUNTER, Thomas C.; FRIEDERICI, Angela D.; SCHRIEFERS, Herbert. Syntactic gender and semantic expectancy: ERPs reveal early autonomy and late interaction. *Journal of Cognitive Neuroscience*, v.12, n.4, pp. 556-568, 2000.

HAGOORT, Peter. How the brain solves the binding problem for language: a neurocomputational model of syntactic processing. *Neuroimage*, v.20, pp. 18-29, 2003.

_____. The syntactic Positive Shift (SPS) as an ERP measure of syntactic processing. *Language and Cognitive Processes*, v. 8, n. 4, pp. 439-483, 1993.

_____. The binding problem for language, and its consequences for the neurocognition of comprehension. In: GIBSON, E. A.; PEARLMUTTER, N. J. (Orgs.). *The processing and acquisition of reference*. Cambridge, MA: MIT Press, pp. 403-436, 2011.

HAHNE, Anja; FRIEDERICI, Angela D.. Electrophysiological Evidence for Two Steps in Syntactic Analysis: Early Automatic and Late Controlled Processes. *Journal of Cognitive Neuroscience*, v.11, n.2, pp. 194–205, 1999.

HEIM, Stefan. Syntactic gender processing in the human brain: A review and a model. *Brain and Language*, v. 106, p. 155-164, 2008.
HERNANDEZ, Arturo E.; KOTZ, Sonja A.; HOFMANN, Juliane; VALENTIN, Vivian V.; DAPRETTO, Mirella; BOOKHEIMER, Susan Y. The neural correlates of grammatical gender decisions in Spanish. *Brain Imaging*, v- 15, n. 4, Abril/2004.

HERNANDEZ, Arturo E.; HOFMANN, Juliane; KOTZ, Sonja A. Age

of acquisition modulates neural activity for both regular and irregular syntactic functions. *NeuroImage*, v. 36, pp.912-923, 2007.

HOCKET, Charles. F. *A course in modern linguistics*. Nova York, The Macmillan Company, 1958.

HOHFELD, Annette. Accessing gramatical gender in german: the impact of gender-marking regularities. *Applied Psycholinguistics*, v. 27, pp. 127-142, 2006.

HOOPER, Joan Bybee. Child Morphology and Morphophonemic change. In: FISIAK, J. (org.) *Historical Morphology. Trends in Linguistics: Studies and Monographs*, v.17, pp. 157-187. The Hague: Mouton, 1980.

JACKENDOFF, Ray. What's in the lexicon.? In: NOOTEBOOM, Sieb.; WEERMAN, Fred.; WIJNEN, Frank.(Orgs). *Storage and Computation in the Language Faculty*, Dordrecht: Kluwer Academic Publishers, pp. 61–92, 2002.

JESCHENIAK, Jörg. D., LEVELT, Willem. J. M. Word frequency effects in speech production: Retrieval of syntactic information and of phonological form. *Journal of Experimental Psychology: Language, Memory and Cognition*, v.20, pp. 824-843, 1994.

JOANISSE, Marc F., SEIDENBERG, Mark. S.. Impairments in verb morphology after braininjury: A connectionist model. *Proceedings of the National Academy of Science USA*, v.96, pp.7592– 7597, 1999.

KAAN, Edith; SWAAB, TamaraY.. Repair, revision, and complexity in syntactic analysis: an electrophysiological differentiation. *Journal of Cognitive Neuroscience*, v.15, pp. 98-110, 2003.

KAAN, Edith. Event-Related potentials and language processing: A Brief overview. *Language and Linguistics Compass*, v. 1, n.6, pp. 571-591, 2007.

KUTAS, Martha; HILLYARD, Steven A.. Reading senseless sentences: brain potentials reflect semantic incongruity. *Science*, v.207, pp. 203–05, 1980.

- KUTAS, Martha; VAN PETTEN, Cyma K.. Psycholinguistics Electrified: Event-related potential investigations. In: TRAXLER, Matthew J.; GERNSBACHER, Morton Ann (Orgs.), *Handbook of Psycholinguistics*, London: Academic Press, pp. 83-143, 1994.
- KUTAS, Marta; FEDERMEIER, Kara D.. Thirty years and counting: Finding meaning in the N400 component of the event related brain potential (ERP), *Annual Review of Psychology*, v.62, pp. 621-647, 2010.
- LAI, Vicky Tzuyin; CURRAN, Tim. ERP evidence for conceptual mappings and comparison processes during the comprehension of conventional and novel metaphors. *Brain and Language*, v.127, n.3, pp.484-496, 2013.
- LAWRENCE, Michael A.. *ez: Easy analysis and visualization of factorial Experiments*, 2013. R package version 4.2-2. <http://CRAN.R-project.org/package=ez>
- LEVELT, Willem.J.M.; PRAAMSTRA, Peter; MEYER, Antje.S.; HELENIUS, Päivi, SALMELIN, Riitta. An MEG study of picture naming. *Journal of Cognitive Neuroscience*, v. 10, pp.553-567, 1998.
- LUCCHESI, Dante. A categoria gramatical de gênero: universais, mudança e crioulização. In: *Razões e Emoções. Miscelânea de estudos em Homenagem a Maria Helena Mira Mateus*, v. 1. Lisboa, Imprensa Nacional-Casa da Moeda, pp. 429-450, 2003.
- MACCULLOCH, Warren., PITTS, Walter. A logical calculus of ideas immanent in nervous activity, *Bulletin of Mathematical Biophysics*, v. 5, pp. 115-133, 1943.
- MACWHINNEY, Brian; LEINBACH, Jared. Implementations are not conceptualisations: Revising the verb learning model. *Cognition*, v. 29, pp. 121-157, 1991.
- MACWHINNEY, Brian. The competition model: The input, the context, and the brain. In: ROBINSON, Peter (Org.). *Cognition and Second Language Instruction*. Cambridge: Cambridge University Press, pp. 69-90, 2001.
- MACWHINNEY, Brian. New directions in the competition model. In.

- TOMASELLO, Michael; SLOBIN, Dan I. (Orgs.). *Beyond Nature-Nurture: Essays in honor of Elizabeth Bates*. Mahwah, NJ: Lawrence Erlbaum, 2005.
- MARGOTTI, Felício Wessling. *Morfologia do Português*. Florianópolis: LLV/CCE/UFSC, 2008.
- MIRKOVIC, Jelena; MACDONALD, Maryellen C.; SEIDENBERG, Mark S.. Where does gender come from? Evidence from a complex inflectional system. *Language and Cognitive Processes*, v. 19, 2004.
- MIRKOVIC, Jelena; SEIDENBERG, Mark S.; JOANISSE, Marc F. Rules vs. Statistics: Insights from a Highly Inflected Language. *Cognitive Science*, v. 35, n.4, pp. 638-681, 2011.
- MOLINARO, Nicola; BARBER, Horacio; CARREIRAS, Manuel. Gramatical Agreement processing in reading: ERP findings and future directions. *Cortex*, v. 47, pp.908-930, 2011.
- MOLINARO, Nicola; KIM, Albert; VESPIGNANI, Francesco; JOB, Remo. Anaphoric agreement violation: An ERP analysis of its interpretation. *Cognition*, v.106, n.2, pp. 963-974, 2008.
- MORGAN-SHORT, Kara; SANZ, Cristina; STEINHAEUER, Karsten; ULLMAN, Michael.T; Second Language Acquisition of Gender Agreement in Explicit and Implicit Training Conditions: An Event-Related Potential Study. *Language Learning*, v.60, n.1, pp. 154-193, 2010.
- MUNTE, Thomas F.; HEINZE, H.J. ERP negativities during syntactic processing of written words, In: HEINZE, H.J., MUNTE, T.F., MANGUN, G.R. (Orgs.). *Cognitive electrophysiology*. Boston: Birkhauser, 1994.
- MÜNTE, Thomas F.; TESSA, Say; CLAHSSEN, Harald; SCHILTZ, Kolja; KUTAS, Marta. Decomposition of morphologically complex words in English: evidence from event-related brain potentials. *Cognitive Brain Research*, v.7, pp.241-253, 1999.
- NAME, Maria Cristina Lobo. *Habilidades perceptuais e linguísticas no processo de aquisição do sistema de gênero no português*. Tese

(Doutorado). Rio de Janeiro: Pontifícia Universidade Católica do Rio de Janeiro, 2002.

NAME, Maria Cristina Lobo; CORREA, Letícia Maria Sicuro. Delimitação Percentual de uma Classe Correspondente à Categoria Funcional D: Evidências da Aquisição do Português. *Revista Fórum Lingüístico*, v.3, nº1, pp.55-88, 2003.

NASCIMENTO, Mauro José Rocha do. Repensando as vogais temáticas nominais a partir da gramática das construções. 2006. 171 f. Tese (Doutorado) - Departamento de Letras Vernáculas, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2006.

NEWMAN, Aaron J.; ULLMAN, Michael T.; PANCHEVA, Roumyana; WALIGURA, Diane L.; NEVILLE, Helen J..An ERP study of regular and irregular English past tense inflection. *NeuroImage*, v.34, pp.435-445, 2007.

OLDFIELD, R. C; WINGFIELD, A. Response latencies in naming objects. *Quarterly Journal of Experimental Psychology*, v.17, pp. 273-281, 1965.

OSTERHOUT, Lee; McKINNON, Richard; BERSICK, Michael; COREY, Vicka. On the language- specificity of the brain response to syntactic anomalies: Is the syntactic positive shift a member of the P300 family? *Journal of Cognitive Neuroscience*, v. 8, pp. 507–526, 1996.

PADOVANNI, Roberto; CALANDRA-BUONAURA, Giovanna; CACCIARI, Cristina; BENUZZI, Francesca; NICHELLI, Paolo. Gramatical Gender in the brain: Evidence from an fMRI study on Italian. *Brain Research Bulletin*, v. 65, pp.301-308, 2005.

PADOVANI, Roberto. Gramatical Gender and morphological transparency in Italian words. Poster presented at theoretical and experimental linguistics, Euroconference on: The syntax of normal and impaired Language, Corinth, Greece, 2002.

PADOVANI, Roberto; CACCIARI, Giovanna. The role of morphological transparency in Italian words recognition, *Giornale Italiano di Psicologia*, v.4, pp.749-771, 2003.

PINKER, Steven. Rules of Language. *Science*, v.253, p. 530-535, 1991.

PRASADA, Sandeep; PINKER, Steven. Generalizations of regular and irregular morphology. *Language and Cognitive Processes*, v.8, n.1, pp. 1-56, 1993.

PINKER, Steven. Words and Rules. *Lingua*, v.106, pp.219-242, 1998.

PINKER, Steven. *Words and rules: the ingredients of Language*. Londres: Weidenfeld and Nicolson, 1999.

PINKER, Steve; ULLMAN, Michael. T. The past and future of the past tense. *Trends in Cognitive Sciences*, v.6, n.11, pp.456-463, 2002.

PLUNKET, Kim; MARCHMAN, Virginia. U-shaped learning and frequency effects in a multi-layered perceptron: Implications for child language acquisition. *Cognition*, v. 38, pp. 43-102, 1991.

PLUNKETT, Kim; MARCHMAN, Virginia. From rote learning to system building: Acquiring verb morphology in children and connectionist nets. *Cognition*, v. 48, pp. 21–69, 1993.

PLUNKETT, Kim; JUOLA, Patrick. A connectionist model of English past tense and Plural Morphology. *Cognitive Science*, v. 23, pp. 463-490, 1999.

PRADO, Elizabeth; ULLMAN, Michael. T. Can imageability help us draw the line between storage and composition? *Journal of Experimental Psychology: Language, Memory and Cognition*, v. 35, n.4, p.849-866, 2009.

PRASADA, Sandeep, PINKER, Steven. Generalizations of regular and irregular morphological patterns. *Language and Cognitive Processes*, 8, 1-56, 1993.

R CORE TEAM. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. Disponível em: URL <http://www.R-project.org/>, 2013.

RABUÑAL, Juan; DORADO, Julián. *Artificial Neural Networks in Real life applications*. Hershey: Idea Group Publishing Inc., 2006.

RADFORD, Andrew. *Syntax: a minimalist introduction*. Cambridge: UK University Press, 1997.

ROCHA, Luiz Carlos de Assis. *Estruturas Morfológicas do Português*. 2ª edição, São Paulo: WMF Martins Fontes, 2008.

RUBENSTEIN, Hebert; GARFIELD, Lonnie; MILUKAN, Jane. Homographic entries in the internal lexicon. *Journal of Verbal Learning & Verbal Behavior*, v.9, pp.487-494, 1970.

RUMELHART, David. E., MCCLELLAND, James. L. On learning the past tenses of English verbs. In. MCCLELLAND, James. L.; RUMELHART, David. E.; PDP RESEARCH GROUP (Orgs.). *Parallel distributed processing: Explorations in the microstructures of cognition*. Cambridge, Massachusetts: Bradford/MIT press, v. 2, pp. 216–271, 1989.

SAID ALI, Manuel. *Gramática histórica da língua Portuguesa*. Rio de Janeiro: Livraria Acadêmica. 1971.

SEIDENBERG, Mark; BRUCK, Maggie. Consistency effects in the generation of past tense morphology. Paper presented at the Thirty-first Meeting of the Psychonomics Society, New Orleans, 1990.

TABAK, Wieke; SCHREUDER, Robert; BAAYEN, R. Harald. Producing inflected verbs: A picture naming study. *The Mental Lexicon*, v.5, n.1, pp. 22-46, 2010.

TAFT, Marcus. Recognition of affixed words and the word frequency effect. *Memory and Cognition*, V.7, pp. 263–272, 1979.

TAFT, Marcus; MEUNIER, Fanny. Lexical Representation of Gender: A Quasiregular Domain. *Journal of Psycholinguistic research*, v. 27, n. 1, pp. 23- 45, 1998.

TUCKER, G. Richard; LAMBERT, Wallace E.; RIGAULT, A. A.; SEGALOWITZ, N. A psychological investigation of French speaker's skill with grammatical gender. *Journal of Verbal Learning and Verbal Behavior*, n. 7, pp. 312-316, 1968.

TUCKER, G. Richard; LAMBERT, Wallace. E.; RIGAULT, A. A.. *The French speaker's skill with grammatical gender: An example of rule-governed behavior*. The Hague: Mouton, 1977.

ULLMAN, Michael. T.; CORKIN, Suzanne; COPPOLA, Marie; HICKOK, Gregory; GROWDON, John H.; KOROSHETZ, Walter J.; PINKER, Steven. A neural dissociation within language: Evidence that the mental dictionary is part of declarative memory, and that grammatical rules are processed by the procedural system. *Journal of Cognitive Neuroscience*, v.9, pp. 266-276, 1997.

ULLMAN, Michael T. Acceptability ratings of regular and irregular past-tense forms: Evidence for a dual-system model of language from word frequency and phonological neighbourhood effects. *Language and Cognitive Processes*, v. 1, n. 14, p. 47-67, 1999.

ULLMAN, Michael T. The declarative/procedural model of lexicon and grammar. *Journal of Psycholinguistic Research*, v.30, n.1, pp. 37-69, 2001a.

ULLMAN, Michael T. A neurocognitive perspective on language: The declarative/procedural model. *Nature Reviews Neuroscience*, v.2, p.717-726, 2001b.

ULLMAN, Michael T.; ESTABROOKE, Ivy V.; STEINHAUER, Karsten; BROVETTO, Claudia; PANCHEVA, Roumyana; OZAWA, Kaori; MORDECAI, Kristen; MAKI, Pauline. Sex differences in the neurocognition of language, *Brain and Language*, v. 83, pp. 141–143, 2002.

ULLMAN, Michael T. Contributions of memory circuits to language: The declarative/procedural model. *Cognition*, v.92, n.1-2, pp. 231-270, 2004.

ULLMAN, Michael T. A cognitive neuroscience perspective on second language acquisition: The declarative/procedural model. In: SANZ, C. (Org). *Adult second language acquisition*. Washington, DC: Georgetown University Press, pp. 141-178, 2005.

ULLMAN, Michael T. Is Broca's area part of a basal ganglia thalamocortical circuit? *Cortex*, v.42, n.4, pp. 480-485, 2006.

VILLALVA, Alina. *Estruturas morfológicas: unidades e hierarquias nas palavras do português*. Lisboa: Fundação Calouste Gulbenkian, 2000.

VAN DER LELY, Heather, K.J., ULLMAN, Michael, T. Past tense morphology in specifically language impaired and normally developing children. *Language and Cognitive Processes*, n. 16 (2/3), 177–217, 2001.

ZANOTTO, Normélio. *Estrutura mórfica da língua portuguesa*. Caxias do Sul: EDUCS, 1986.

APÊNDICE A: Termo de consentimento

Universidade Federal de Santa Catarina
Centro de Comunicação e Expressão -
Departamento de Línguas Vernáculas
Programa de Pós Graduação em Linguística
LabLing– Laboratório da Linguagem e Processos Cognitivos

PROJETO DE PESQUISA: INVESTIGANDO A NEUROCOGNIÇÃO
DA MORFOLOGIA DE GÊNERO DO PB COMO L1 E L2

Nome _____ do
Participante: _____

Pais ou responsáveis (no caso de menor de 18 anos):

Endereço:

Cidade:

Estado: _____ CEP: _____ Telefone:

(____) _____ RG: _____ CPF:

nascimento: ____/____/____ Data da coleta:

____/____/____

Nome da Pesquisadora Principal: Mailce Borges Mota Outros
pesquisadores: Natália Carolina Alencar de Resende

Eu, Natália Carolina Alencar de Resende, CPF: 053.939.876-41, RG: MG-5 659-650, aluna de doutorado do Programa de Pós-graduação em Linguística da Universidade Federal de Santa Catarina, tenho como objetivo desenvolver um estudo sobre a neurocognição da morfologia de gênero do português do Brasil como língua materna e segunda língua, requisito parcial para a obtenção do título de doutor(a) em Linguística. Para tanto, o estudo contará com a participação de voluntários na execução de dois tipos de tarefas: uma tarefa comportamental e uma tarefa para captação de sinais eletrofisiológicos.

Procedimentos:

Tarefa comportamental: Os voluntários devem relizar a concordância de gênero ao pressionar as teclas “a” ou “o” do teclado do computador em estímulos linguísticos apresentados visualmente no monitor de um computador. Esta tarefa terá duração máxima de 30min.

Tarefa para captação de sinais eletrofisiológicos: Os voluntários serão submetidos à colocação de eletrodos sobre o couro cabeludo cuja fixação será realizada através de uma pasta condutora. Após a colocação dos eletrodos, os voluntários devem ler, silenciosamente, as sentenças apresentadas no monitor de um computador e julgá-las como corretas ou incorretas. Para esse julgamento, os voluntários devem pressionar as teclas “1” para sentenças corretas e “0” para as sentenças incorretas. Essa tarefa terá duração média de 1h30min no total.

Benefícios:

A participação nos experimentos será voluntária e não resultará em qualquer benefício direto para o voluntário.

Riscos:

Nenhuma das tarefas fornece risco físico ou moral para os participantes. O experimento para captação de sinais eletrofisiológicos dos participantes não fornece risco de choques elétricos.

Desconforto:

A colocação dos eletrodos é um processo lento (em torno de 45min/1h) o que pode acarretar certo desconforto para os participantes. Durante a colocação, ofereceremos filmes, livros, revistas para minimizar esse desconforto. Nós certificaremos, também, de que o ambiente do laboratório LabLing oferece condições satisfatórias para a execução das tarefas, como: iluminação do ambiente, temperatura condicionada, posicionamento adequado do monitor do computador (de acordo com a altura de cada participante) e cadeiras confortáveis.

Direitos dos participantes:

Os participantes podem desistir do estudo a qualquer momento sem sofrer nenhum prejuízo e têm o direito de acesso, em qualquer etapa do estudo,

sobre qualquer esclarecimento de eventuais dúvidas. Tendo qualquer dúvida sobre a pesquisa, os participantes podem entrar em contato com Natália Resende através do email natcarol2000@gmail.com, telefone (48) 96202780, ou com a professora Dra. Mailce Borges Mota através do email mailcemota54@gmail.com, telefone (48) 3721-9288, ou no prédio do Centro de Comunicação e Expressão – CCE, bloco B, sala 101, Universidade Federal de Santa Catarina – UFSC.

Compensação financeira:

Não existirão despesas pessoais ou compensações financeiras relacionadas à participação no estudo. Qualquer despesa adicional será absorvida pelo orçamento da pesquisa.

Utilização dos dados:

Eu me comprometo a utilizar os dados coletados somente para pesquisa e os resultados serão apresentados na forma de Tese de Doutorado e veiculados através de artigos científicos em revistas especializadas e/ou em encontros científicos e congressos, após a aprovação do estudo pelo Programa de Pós Graduação em Linguística. Garanto que a identidade dos participantes deste estudo será mantida em sigilo e, de modo algum, será revelada.

Termo de consentimento livre e esclarecido:

Eu compreendo meus direitos como um sujeito de pesquisa e voluntariamente concordo em participar deste estudo e em ceder meus dados para pesquisa. Compreendo sobre o que, como e porque este estudo está sendo feito. Compreendo, também, os objetivos do estudo, os procedimentos a serem realizados, possíveis desconfortos e riscos, as garantias de confidencialidade e de esclarecimentos permanentes. Receberei uma cópia assinada deste formulário de consentimento.

Assinatura do sujeito:

participante Assinatura do pesquisador:

APÊNDICE B: Estímulos Experimento 1

Estímulos	Nº de grafemas	Frequência	Categoria
lápis	3	2119	Opaco
saúde	5	320	Opaco
clave	5	1	Opaco
coldre	6	1	Opaco
covil	5	1	Opaco
dogma	5	3	Opaco
filme	5	597	Opaco
gente	5	822	Opaco
lei	3	572	Opaco
limiar	6	6	Opaco
mês	3	905	Opaco
noite	5	616	Opaco
ode	3	1	Opaco
país	4	1461	Opaco
time	6	385	Opaco
plebe	5	1	Opaco
prole	5	3	Opaco
prelo	5	2	Regular
ata	3	10	Regular
átrio	5	1	Regular

bojo	4	5	Regular
casa	4	1062	Regular
casta	5	7	Regular
fleuma	6	1	Regular
hora	4	480	Regular
hulha	5	1	Regular
jogo	4	689	Regular
jugo	4	2	Regular
livro	5	648	Regular
preço	5	407	Regular
tempo	5	1515	Regular
vida	4	1059	Regular
Pseudopalavras			
plefe			
clage			
bilme			
faís			
movil			
noife			
sogma			
laúde			
dota			
visma			
pesmo			

lespo	
ferna	
tirca	
fospo	
narbo	
itade	
sagem	
jonte	
lidez	
lipré	
opagem	
idote	
bonte	

APÊNDICE C: Estímulos Experimento 2

Estímulos	Categoria	Frequência	Nº de grafemas
Cisão	Irregular	7	5
Cinema	Irregular	401	6
Teorema	Irregular	4	7
Sistema	Irregular	731	7
Remissão	Irregular	3	8
Lema	Irregular	11	4
Região	Irregular	642	6
Mão	Irregular	563	3
Unção	Irregular	1	5
Programa	Irregular	958	8
Edema	Irregular	3	5
Recensão	Irregular	1	8
Tema	Irregular	352	4
Diagrama	Irregular	5	8
População	Irregular	347	9
Razão	Irregular	352	5
Nuance	Opaco	4	6
Filme	Opaco	883	5
Gente	Opaco	815	5
Nome	Opaco	869	4
Frente	Opaco	578	6

Plebe	Opaco	1	5
Covil	Opaco	1	5
Ventre	Opaco	13	6
Alpendre	Opaco	1	8
Morte	Opaco	486	5
Noite	Opaco	669	5
Prole	Opaco	3	5
Time	Opaco	652	4
Papel	Opaco	468	5
Ode	Opaco	2	3
Âmbar	Opaco	1	5
Clamor	Transparente	6	5
Valor	Transparente	701	5
Setor	Transparente	670	5
Clivagem	Transparente	1	8
Cavidade	Transparente	6	8
Imagem	Transparente	504	6
Reator	Transparente	5	6
Atividade	Transparente	379	9
Fulgor	Transparente	2	6
Amor	Transparente	297	4
Viagem	Transparente	272	6
Verdade	Transparente	487	7
Aragem	Transparente	1	6

Rubor	Transparente	2	5
Fator	Transparente	165	5
Acuidade	Transparente	1	8
Paleta	Regular	1	6
História	Regular	857	8
Carro	Regular	679	5
Fava	Regular	3	4
Tomo	Regular	5	4
Sela	Regular	2	4
Acordo	Regular	723	6
Fardo	Regular	5	5
Rixa	Regular	3	4
Obra	Regular	722	4
Gnomo	Regular	1	5
Limbo	Regular	3	5
Número	Regular	847	6
Prelo	Regular	2	5
Tromba	Regular	1	6
Área	Regular	873	4
Lado	Regular	847	4
Família	Regular	571	7
Resultado	Regular	632	9
Baia	Regular	1	4
Pesquisa	Regular	662	8

Processo	Regular	800	8
Bojo	Regular	5	4
Ideia	Regular	671	5
Pompa	Regular	5	5
Jugo	Regular	2	4
Cultura	Regular	439	7
Período	Regular	502	7
Átrio	Regular	1	5
Entulho	Regular	5	7
Crina	Regular	1	5
Palavra	Regular	476	7
Economia	Regular	494	8
Futuro	Regular	431	6
Salário	Regular	522	7
Água	Regular	421	4
Política	Regular	529	8
Negócio	Regular	341	7
Crivo	Regular	4	5
Linha	Regular	484	5
Obelisco	Regular	2	8
Prazo	Regular	532	5
Recurso	Regular	629	7
Fisco	Regular	10	5
Lousa	Regular	2	5

Corja	Regular	1	5
Moita	Regular	4	5
Calha	Regular	1	5

APÊNDICE D: Estímulos Experimento 3

Estímulos	Categoria	Nº de grafemas	Terminação	Dificuldade
Famida	Reg	6	-a	Difícil
Damume	Transp/Masc	6	-ume	Difícil
Galodor	Transp/Masc	7	-or	Difícil
Cetafo	Reg	6	-o	Difícil
Cetade	Transp/fem	6	-ade	fácil
Maragem	Transp/fem	7	-agem	fácil
Afenção	Transp/Irr	7	-ão	fácil
Voragem	Transp/fem	7	-agem	fácil
Ronstrução	Transp/Irr	10	-ão	fácil
Lontade	Transp/fem	7	-ade	fácil
Titema	Transp/Irr	6	-ema	Difícil
Tomento	Reg	7	-o	fácil
Imade	Transp/fem	5	-ade	fácil
Evema	Transp/Irr	5	-ema	fácil
Fensador	Transp/Masc	8	-or	fácil
Decagem	Transp/fem	7	-agem	Difícil

Fotícia	Reg	7	-a	fácil
Eferta	Reg	6	-a	fácil
Vomagem	Transp/fem	7	-agem	Difícil
Docotão	Transp/Irr	7	-ão	Difícil
Sor	Transp/Mas c	3	-or	fácil
Cetoda	Reg	6	-a	Difícil
Damibo	Reg	6	-o	Difícil
Secade	Transp/fem	6	-ade	Difícil
Liume	Transp/Mas c	5	-ume	fácil
Pobedor	Transp/Mas c	7	-or	Difícil
Merfume	Transp/Mas c	7	-ume	fácil
Bamília	Reg	7	-a	fácil
Cilema	Transp/Irr	6	-ema	fácil
Priação	Transp/Irr	7	-ão	fácil
Fomadão	Transp/Irr	7	-ão	Difícil
Sisfema	Transp/Irr	7	-ema	fácil
Ligume	Transp/Mas c	6	-ume	fácil
Nor	Transp/Mas c	3	-or	fácil
Vomade	Transp/fem	6	-ade	Difícil

Muturo	Reg	6	-o	fácil
Cedema	Transp/Irr	6	-ema	Difícil
Tetadão	Transp/Irr	7	-ão	Difícil
Cecuto	Reg	6	-o	Difícil
Vabade	Transp/fem	6	-ade	Difícil
Datume	Transp/Mas c	6	-ume	Difícil
Detuca	Reg	6	-a	Difícil
Favema	Transp/Irr	6	-ema	Difícil
Fafeco	Reg	6	-o	Difícil
Fanodor	Transp/Mas c	7	-or	Difícil
Vostume	Transp/Mas c	7	-ume	fácil
Cabagem	Transp/fem	7	-agem	Difícil
Tantagem	Transp/fem	8	-agem	fácil

APÊNDICE E: Inputs apresentados à rede neural

Palavras	Frequência de entrada no arquivo de treino
Cultura	45
Linha	49
Economia	50
Palavra	48
Política	64
Idea	67
Área	87
História	85
Família	57
Água	72
Pesquisa	62
Resultado	63
Carro	68
Acordo	72
Processo	80

Lado	84
Número	84
Futuro	43
Período	50
Salário	52
Recurso	63
Negócio	34
Setor	67
Valor	70
Fator	16
Amor	29
Viagem	50
Imagem	56
Atividade	37
Mão	56
Região	64
Razão	35
População	37
Sistema	73
Cinema	40

Problema	103
Nome	55
Filme	88
Noite	66
Time	65
Total	2386

APÊNDICE F: Estímulos experimento eletrofisiológico

Códigos de trigger de cada condição

Códigos de trigger	Significado
111	Violada/adjetivo/regular
112	Violada/adjetivo/irregular
121	Violada/substantivo/regular
122	Violada/substantivo/irregular
221	Não Violada/substantivo/regular
211	Não Violada/adjetivo/regular
222	Não Violada/substantivo/irregular
212	Não violada/adjetivo/irregular

Perguntas apresentadas aos sujeitos para controle de atenção

Bloco Nº	Item Nº	Perguntas
1	1	A ligação foi gravada?
1	2	O ritmo das aulas era lento?
1	3	A injeção casou uma dor na cabeça?
1	4	O paciente terminou o tratamento de saúde?
1	5	Os passageiros perderam o horário do voo?
1	6	Os surdos esqueceram a linguagem de sinais?
2	1	A farmácia dispensou a receita do remédio?
2	2	O código genético do DNA foi descoberto pelos cientistas?
2	3	O diretor escondeu a situação crítica da escola?
2	4	A conversou aconteceu em um evento?
2	5	Os passageiros perderam o ônibus?
2	6	A capacidade do sistema estava descrita no manual?
3	1	Os livros tinham uma capa dura?
3	2	O show foi de música sertaneja?
3	3	O show iniciou no horário marcado?
3	4	O curso discutiu filosofia?
3	5	A saúde recebeu investimento do governo?
3	6	Houve chuva no oeste do estado?
4	1	O professor calculou a nota dos alunos na prova?

4	2	O estado apostou no turismo rural?
4	3	O estudantes tiveram uma imagem positiva da univerisidade?
4	4	o time sofreu mudanças?
4	5	O reunião aconteceu no domingo?
4	6	A agência terminou a seleção dos candidatos?

Conjunto de frases apresentadas aos sujeitos

pre_target	target	post_target	Códigos	Frequência por milhão
O/documentário/mostrou/a/obra	clássico	do/pintor	111	343
O/curso/cobriu/a/história	literário	do/Brasil	111	322
O/partido/criticou/a/estrutura	político	do/grupo	111	440
O/biólogo/filtrou/a/água	salgado	da/solução	111	216
A/lei/aprovou/a/pesquisa	acadêmico	na/escola	111	495
O/banco/financiou/a/empresa	estrangeiro	no/país	111	399
O/acordo/abrangeu/a/área	jurídico	da/região	111	328
Os/imigrantes/adotaram/a/prática	religioso	da/comunidade	111	216
Os/reforços/melhoraram/a/média	baixo	das/notas	111	500
Os/operários/instalaram/a/estrutura	metálico	do/prédio	111	364
O/pesquisador/participou/da/mesa	redondo	da/conferência	111	257
O/professor/definiu/a/área	temático	do/estudo	111	283
O/governo/construiu/a/escola	básico	da/comunidade	111	500
O/quadro/retratava/a/figura	feminino	do/período	111	283
O/juiz/atribuiu/a/pena	máximo	ao/réu	111	290
O/oriente/rejeitou/	moderno	do/ocident	111	265

a/cultura		e		
A/escola/ofereceu/apoio	pedagógica	aos/país	111	186
O/curso/formou/um/corpo	técnica	de/qualidade	111	447
As/crianças/sentaram/no/banco	traseira	do/automóvel	111	369
Os/visitantes/encontraram/o/ponto	turístico	da/cidade	111	246
As/medidas/afetaram/o/comércio	externa	do/país	111	277
O/professor/citou/um/caso	extrema	do/problema	111	190
O/banco/aumentou/o/prazo	mínima	dos/pagamentos	111	571
O/empregado/assinou/o/contrato	temporária	de/emprego	111	425
A/redação/tornou-se/um/processo	complexa	para/adolescentes	111	184
O/casal/deu/um/passo	decisiva	na/vida	111	673
O/designer/enviou/o/projeto	gráfica	por/e-mail	111	426
A/plateia/aplaudiu/o/discurso	jurídica	do/governador	111	284
O/médico/relatou/o/estado	clínica	do/paciente	111	287
A/pesquisa/realizou/um/estudo	descritiva	da/população	111	269
As/crianças/bricaram/no/lado/	externa	da/casa	111	322
O/voto/representou/o/exercício	plena	da/cidadania	111	352
O/cientista/defendeu/a/verdade/	absoluto	da/ciência	112	452

Os/alunos/são/a/parte	ativo	da/universidade	112	293
O/professor/corrigiu/o	resposta	da/aluna	121	468
O/capítulo/descreveu/a/abordagem	teórico	da/pesquisa	112	420
O/subsídio/beneficiou/a/classe	baixo	da/escola	112	494
O/Brasil/cresceu/no	década	da/ditadura	121	281
Os/pesquisadores/discutiram/a/questão	ético	do/estudo	112	402
O/modelo/apresentou/uma/ordem	hierárquico	dos/elementos	112	212
Os/especialistas/criticaram/a/base	científico	do/modelo	112	202
O/diretor/reestruturou/a/saúde	financeiro	da/empresa	112	314
O/prefeito/reconheceu/a/ação	rápido	do/presidente	112	309
O/produtor/informou/o	renda	da/família	121	270
A/faculdade/divulgou/o	lista	de/aprovados	121	273
A/festa/aconteceu/numa/cidade	pequeno	do/estado	112	303
A/faculdade/inaugurou/uma/sede	próprio	no/centro	112	485
O/prefeito/reformou/a/região	urbano	da/cidade	112	221
O/condomínio/lançou/uma	espaço	para/crianças	121	293
O/público/prestigiou/o/filme	estrangeira	do/festival	112	257
A/população/mobil	inteira	com/passe	112	490

izou/o/país		atas		
A/prova/atingiu/o/ grau	máxima	de/dificuld ade	112	261
O/funcionário/ocu pou/a	cargo	do/diretor	121	190
O/governo/organiz ou/um/sistema	complex a	de/transaç ões	112	451
O/professor/defini u/um/tema	específic a	para/discu ssão	112	149
A/enfermeira/pedi u/o/exame	médica	do/pacient e	112	421
A/reunião/discutiu/ a	assunto	da/pauta	121	327
O/ex- técnico/treinou/o/ti me	mineira	no/campe onato	112	492
Os/cortes/represent aram/um/fator	positiva	no/orçame nto	112	341
A/prefeitura/refor mou/o/transporte	urbana	da/cidade	112	365
Os/diretores/infor maram/o/nome	artística	dos/atores	112	470
Os/jornalistas/denu nciaram/o/crime	hedionda	na/TV	112	345
A/corrupção/criou/ um/problema	política	no/Brasil	112	473
A/guerra/acontece u/na	século	do/ouro	121	267
O/governo/investiu /no	educação	do/país	122	265
A/banda/reviveu/o	época	dos/Beatle s	121	261
A/imprensa/divulg ou/o	guerra	na/Bósnia	121	311

As/fotos/revelaram/o	natureza	do/caso	121	195
Os/remédios/curaram/o	doença	dos/pacientes	121	417
Os/alunos/observaram/o	maneira	de/escrever	121	355
A/agência/apresentou/um	proposta	de/campanha	121	740
O/governador/implantou/um	universidade	no/estado	122	247
O/governador/impulsionou/o	economia	da/região	121	221
As/vítimas/acreditaram/no	justiça	do/julgamento	121	360
A/testemunha/solicitou/o	presença	do/advogado	121	173
O/casal/planejou/um	viagem	de/férias	122	255
A/caneta/manchou/o	folha	de/tinta	121	240
Os/jogadores/discutiram/a	futuro	do/time	121	415
Os/noivos/escolheram/o	ordem	das/músicas	122	281
As/crianças/aprenderam/um	palavra	por/dia	121	330
A/polícia/registrou/a	acidente	de/carro	122	249
Os/políticos/fretaram/uma	avião	da/Gol	121	146
A/universidade/construiu/uma	centro	de/convenções	121	507
Os/médicos/avaliaram/o	efeito	da/medicação	121	629
A/revista/aceitou/a	artigo	do/estudante	121	388

		te		
O/helicóptero/pousou/na	campo	da/fazenda	121	241
A/entrevista/aconteceu/na	prédio	da/reitoria	121	139
O/shampoo/aumentou/a	volume	do/cabelo	122	233
O/funcionário/conquistou/a	emprego	dos/sonhos	121	437
O/prefeito/reformou/a	hospital	da/cidade	122	241
Os/filósofos/questionaram/a	sentido	da/vida	121	456
Os/alunos/receberam/a	texto	da/palestra	121	489
O/banco/concedeu/a	crédito	ao/agricultor	121	326
O/orador/anunciou/a	início	da/sessão	121	514
A/neve/caiu/na	norte	do/país	122	262
O/pesquisador/apresentou/a	objetivo	do/estudo	121	408
As/atividades/incrementaram/o	aprendizagem	dos/estudantes	122	296
As/crianças/concluíram/o	atividade	do/dia	122	354
Os/atletas/controlaram/os	fontes	de/energia	122	199
O/vapor/representa/a/fase	gasoso	da/água	112	621
O/pesquisador/realizou/uma/análise	cuidadoso	dos/dados	112	250
A/secretária/assumiu/o	função	do/chefe	122	305

A/pesquisa/conside rou/o	idade	dos/partici pantes	122	208
O/sol/iluminou/o	tarde	de/oming o	122	305
O/deputado/relatou /a/crise	interno	do/congre sso	112	265
O/estudo/apresento u/um	série	de/experi mentos	122	271
Os/jornais/denunci aram/o	situação	dos/bairro s	122	194
O/hospital/reformo u/o	unidade	de/atendi mento	122	222
A/revista/mostrou/ a/relação	amoroso	do/casal	112	293
O/economista/com entou/o	crise	da/Europa	122	304
Os/professores/tom aram/o	frente	da/situaçã o	122	460
Os/trabalhadores/i niciaram/o	greve	da/fábrica	122	229
Os/pedagogos/orga nizaram/o/program a	educativa	da/escola	112	366
O/pedreiro/compro u/a	material	de/constru ção	122	267
A/reportagem/anun ciou/a	morte	do/cantor	122	462
Os/livros/estimular am/a	prazer	da/leitura	122	554
O/RJ/tornou-se/o	cidade	do/samba	122	504
Os/exercícios/ajud aram/na	controle	do/diabete s	122	246
Os/alunos/aprender am/um/idioma	estrangei ra	no/curso	112	399
O/motorista/passou	teste	de/direção	122	160

/na				
Os/jovens/encontraram/um/local	adequada	para/festas	112	460
Os/professores/utilizaram/a	computador	das/salas	122	191
A/notícia/apareceu/na	jornal	do/bairro	122	448
Os/resultados/revelaram/a	nível	do/desempenho	122	345
O/chefe/elogiou/a	ambiente	de/trabalho	122	199
As/análises/revelaram/a	índice	de/chuvas	122	346
Os/turistas/visitaram/a	parque	da/cidade.	122	345
Os/alunos/iniciaram/o/debate	teórica	do/tema	112	254
A/editora/lançou/a/capa	dura	do/livro	211	335
A/nutricionista/receitou/uma/dieta	líquida	ao/paciente	211	255
O/professor/calculou/a/nota	média	da/prova	211	472
O/Brasil/teve/uma/década	perdida	em/desenvolvimento	211	287
O/cantor/concedeu/uma/entrevista	exclusiva	à/TV	211	274
A/comunidade/exigiu/uma/resposta	rápida	do/prefeito	211	362
A/brincadeira/viou/uma/coisa	séria	para/adultos	211	398
A/secretária/gravou/a/conversa	telefônica	da/negociação	211	210
A/direção/concedeu/carta	branca	à/equipe	211	248

O/padre/falou/da/v ida	eterna	aos/fiéis	211	308
O/banco/ajustou/a/ taxa	máxima	dos/juros	211	291
O/show/contou/co m/música	sertaneja	no/repertó rio	211	314
Os/agricultores/rep ortaram/uma/chuva	ácida	na/plantaç ão	211	221
O/experimento/ana lisou/uma/amostra	aleatória	da/popula ção	211	254
A/farmácia/exigiu/ a/receita	médica	do/remédi o	211	289
O/enfermeiro/traba lhou/com/medicina	preventiv a	no/hospita l	211	237
O/aeroporto/contro lou/o/tráfego	aéreo	na/capital	211	240
A/planilha/informo u/o/lucro	bruto	do/mês	211	158
O/sindicato/lutou/p elo/acordo	coletivo	da/categor ia	211	239
A/lei/proibiu/o/trab alho	noturno	dos/menor es	211	422
A/doença/estava/e m/estágio	avançado	no/pacient e	211	400
Os/filhos/herdaram /o/sucesso	econômi co	dos/pais	211	440
A/Inglaterra/aboliu /o/tráfico	negreiro	no/Brasil	211	322
Os/alunos/frequent aram/o/turno	noturno	da/escola	211	336
O/estado/apostou/n o/turismo	ecológic o	nas/serras	211	492
O/carro/apresentou /um/consumo	excessiv o	de/combus tível	211	538
Os/cientistas/desco	genético	do/DNA	211	450

briram/o/código				
As/aulas/começaram/em/ritmo	lento	nas/escolas	211	291
O/negócio/sofreu/um/risco	elevado	de/falência	211	236
Os/bares/concederam/o/aceso	gratuito	à/internet	211	522
A/bola/percorreu/o/caminho	inverso	do/gol	211	422
A/gráfica/imprimiu/em/tamanho	pequeno	as/fotos	211	316
Os/padres/receberam/a/comunidade	cristã	na/igreja	212	209
O/mirante/forneceu/uma/visão	panorâmica	da/cidade	212	392
O/governador/amplicou/a/rede	privada	de/ensino	212	338
O/promotor/pediu/a/prisão	temporária	do/réu	212	438
A/escola/implantou/a/eleição	direta	de/diretores	212	294
O/estudo/analizou/a/fonte	primária	dos/dados	212	554
Os/fogos/antecederam/a/contagem	regressiva	da/virada	212	259
O/prefeito/controleou/a/gestão	urbana	do/minicípio	212	422
As/chuvas/fecharam/a/ponte	aérea	entre/aeroportos	212	312
A/injeção/causou/uma/dor	aguda	no/braço	212	472
O/diretor/denunciou/a/situação	crítica	da/escola	212	253
Os/convidados/dançaram/a/noite	inteira	na/festa	212	375

Os/pintores/escolheram/a/cor	amarela	das/paredes	212	412
O/padre/falou/da/vontade	divina	na/missa	212	303
A/igreja/criticou/a/posição	estratégica	do/bispo	212	566
Os/estudantes/tiveram/uma/imagem	negativa	da/universidade	212	515
A/indústria/viveciou/um/milagre	econômico	no/período	212	241
As/crianças/rasgaram/o/papel	higiênico	do/banheiro	212	349
O/senado/discutiu/a/razão	prática	da/lei	212	310
Os/personagens/encarnaram/o/amor	romântico	do/livro	212	384
A/companhia/instalou/a/luz	elétrica	do/bairro	212	252
O/diretor/acrescentou/um/humor	negro	à/peça	212	229
O/técnico/mudou/o/esquema	tático	do/time	212	398
A/etiqueta/indicava/o/valor	unitário	dos/produtos	212	399
O/sociólogo/estudou/a/sociedade	burguesa	do/Brasil	212	307
Os/atletas/agradeceram/o/calor	humano	da/torcida	212	233
O/livro/criticou/o/regime	repúblicano	do/Brasil	212	267
Os/jornais/divulgaram/o/empate	técnico	dos/clubes	212	417
O/técnico/consertou/o/motor	elétrico	das/máquinas	212	336
A/crítica/elogiou/o	histórico	do/escritor	212	428

/romance				
O/piloto/manteve/a /velocidade	máxima	do/carro	212	809
O/tema/ocupou/um /lugar	privilegia do	no/congre sso	212	182
Os/cantores/atrasar am/a	abertura	do/show	221	255
O/juiz/iniciou/o	julgamen to	dos/proces sos	221	148
O/mapa/garantiu/a	seguranç a	da/viagem	221	343
O/filme/retratou/a	violência	das/guerra s	221	214
Os/estudantes/pro moveram/uma	festa	na/escola	221	413
A/conversa/aconte ceu/na	saída	do/evento	221	215
Os/filhos/cultivara m/a	terra	dos/pais	221	366
O/paciente/interro mpeu/o	tratament o	de/saúde	221	181
Os/devedores/salda ram/a	dívida	no/banco	221	292
Os/empresários/co mpraram/uma	fazenda	de/gado	221	406
A/vaidade/sustenta /a	indústria	de/cosméti cos	221	366
O/artigo/descreveu /os	instrume ntos	de/pesquis a	221	182
O/discurso/acontec eu/na	câmara	de/veread ores	221	447
Os/passageiros/con feriram/o	horário	do/voo	221	206
O/computador/sub stituiu/a	máquina	de/escreve r	221	223

O/técnico/realizou/uma	mudança	no/time	221	415
A/família/frequentava/a	igreja	do/bairro	221	165
A/marca/investiu/no	lançamento	da/coleção	221	226
O/acidente/aconteceu/na	linha	de/trem	221	467
O/curso/introduziu/o	pensamento	dos/filósofos	221	478
Os/filhos/seguiram/o	exemplo	dos/pais	221	312
Os/diretores/estabeleceram/a	faixa	de/salários	221	218
Os/filhos/ampliaram/o	negócio	da/família	221	318
O/prefeito/elogiou/a	polícia	da/cidade	221	463
Os/apresentadores/anunciaram/as	categorias	dos/prêmios	221	416
As/palestras/abriram/o	congresso	de/medicina	221	246
O/advogado/esqueceu/o	documento	no/escritório	221	218
O/programa/suspendeu/o	pagamento	das/férias	221	746
Os/candidatos/rebateram/a	crítica	da/oposição	221	382
O/governo/planejou/o	investimento	na/saúde	221	341
A/árvore/cresceu/no	período	da/primavera	221	400
A/publicidade/impulsionou/o	sucesso	dos/artistas	221	420
O/festival/prestígio	cinema	do/Brasil	222	188

u/o				
Os/arquitetos/acompanharam/a	construção	das/obras	222	286
Os/surdos/aprenderam/a	linguagem	de/sinais	222	167
A/instituição/entendeu/a	necessidade	dos/membros	222	283
Os/índios/demarcaram/o	limite	da/tribo	222	203
Os/trabalhadores/receberam/a	metade	das/férias	222	168
O/estudo/descreveu/a	população	do/município	222	459
A/pesquisa/investigou/a	realidade	da/comunidade	222	214
O/professor/chamou/a	atenção	dos/alunos	222	350
Os/candidatos/enfrentaram/a	dificuldade	das/provas	222	357
Os/passageiros/pegaram/o	ônibus	no/terminal	222	209
Os/clientes/aproveitaram/a	vantagem	das/ofertas	222	183
O/vereador/criticou/a	decisão	do/prefeito	222	336
Os/funcionários/tomaram/a	direção	da/empresa	222	440
Os/passageiros/compraram/a	passagem	do/ônibus	222	459
O/cliente/anotou/o	telefone	da/loja	222	266
Os/funcionários/fizeram/a	avaliação	do/mês	222	188
Os/atletas/treinaram/no	clube	do/município	222	447

A/chuva/passou/no	oeste	do/estado	222	161
A/auditoria/conferiu/a	qualidade	dos/produtos	222	419
A/greve/causou/uma	confusão	nas/ruas	222	465
O/gerente/contratou/o	pessoal	da/limpeza	222	497
O/município/inaugurou/o	restaurante	do/povo	222	249
Os/diretores/organizaram/uma	reunião	no/sábado	222	175
O/técnico/participou/da	comissão	de/arbitragem	222	598
A/empresa/transformou/o	hotel	em/cassino	222	183
A/agência/começou/a	seleção	dos/candidatos	222	334
O/orçamento/reduziu/o	total	dos/gastos	222	413
O/manual/descreveu/a	capacidade	do/sistema	222	283
Os/turistas/compraram/os	dólares	no/banco	222	293
Os/boatos/desgastaram/a	imagem	dos/atores	222	493
O/aparelho/deu/o	sinal	de/início	222	311

APÊNDICE G: Termo de consentimento (EEG)

Max Planck Institute for Psycholinguistics, 2014

Pesquisador (a): Natália Resende, Daniel Achenson **Email:**
natalia.deresende@mpi.nl; Dan.Achenson@mpi.nl

1. Introdução

Você se apresentou para a participação em um estudo de EEG que tem como objetivo entender a produção da linguagem. Sua participação neste estudo é voluntária. Isso significa que você pode desistir de cooperar com esta investigação a qualquer momento. Leia este termo cuidadosamente e não deixe de fazer perguntas caso o texto não esteja claro. O pesquisador ficará feliz em responder qualquer pergunta.

2. Acontecimentos durante a sessão experimental

Como você decidiu participar da presente pesquisa, nós começaremos com a colocação dos eletrodos em uma espécie de touca de banho que colocaremos, a seguir, em sua cabeça. Cada eletrodo é preenchido com um gel. Durante a preparação, sua cabeça será medida e a resistência de cada um dos eletrodos será verificada para que possamos obter um bom sinal de EEG. Depois da preparação, começaremos o experimento.

Para a captação dos sinais de EEG, é muito importante que você tente não se movimentar, ou se movimentar o mínimo possível. É muito importante, também, que você preste bastante atenção na execução da tarefa. Depois do término do experimento, a touca de eletrodos é removida de sua cabeça e você poderá lavar seu cabelo.

Durante o experimento, você deve ler as frases que aparecem na tela do computador com bastante atenção e responder SIM ou NÃO para as perguntas sobre o conteúdo das frases apresentadas. Você responderá as perguntas por meio dos botões Verde e Vermelho da caixa de respostas. O experimento consiste em 1 sessão que dura por volta de 2 horas. Para participar deste experimento, você receberá €10.

3. Coincidências

O experimento não é um exame médico. Os pesquisadores não são treinados para a análise dos resultados sob uma perspectiva médica. No entanto, em casos raros, os resultados podem revelar que você deve consultar um médico. Se este for o caso, o pesquisador irá notificá-lo. Se você não quer ser notificado a este respeito, sua participação neste experimento não é possível.

4. Informação

Para perguntas relacionadas a este estudo, por favor, entre em contato com Natália Resende (email: natalia.deresende@mpi.nl ou natcarol2000@gmail.com) Quando a coleta dos dados terminar e os resultados estiverem disponíveis, se você desejar, é possível pedir um resumo do estudo.

5. Declaração de participação

“Eu declaro que fui informado com clareza da natureza e métodos da pesquisa, como descrito no texto acima. Minhas perguntas foram respondidas satisfatoriamente. Minha participação neste estudo é voluntária. Eu gostaria de ser informado se os resultados do presente estudo são relevantes para fins de ajuda médica.

Eu entendo que tenho o direito de desistir do experimento a qualquer momento.

Se os meus resultados da pesquisa forem usados para fins de publicação ou disponíveis ao público de qualquer outra forma, eles serão divulgados anonimamente. Meus dados pessoais não serão acessados por terceiros sem que haja meu explícito consentimento.”

Nome em CAIXA ALTA:

Assinatura:

Nome do pesquisador:

Assinatura do pesquisador:

Data:

Local: Nijmegen

APÊNDICE H: Debriefing

Estudo: Representação e processamento do gênero gramatical do Português do Brasil

Este experimento tem como principal objetivo investigar o processamento cerebral do gênero gramatical do português, mais especificamente, do português do Brasil. Em português, atribuímos gênero (feminino ou masculino) a todas as palavras. Por exemplo, em português, a palavra *mesa* é feminina, e a palavra *carro* é masculina. Grande parte das palavras que terminam em *-a* são femininas, e que terminam em *-o* são masculinas. Essas palavras são consideradas transparentes quanto ao seu gênero gramatical, pois podemos predizer se a palavra é feminina ou masculina com base em sua forma (terminação). No entanto, um grande conjunto de palavras apresenta outras terminações. Por exemplo, *noite* é uma palavra feminina com terminação em *-e*, *leite* é uma palavra masculina e termina em *-e* também. Além disso, há muitas palavras masculinas que terminam em *-a* como *problema*, *sistema*. Outras terminam em *-o* e são femininas como *moto* e *foto*. Essas palavras são consideradas opacas quanto ao seu gênero gramatical, pois **não** é possível predizer se a palavra é feminina ou masculina com base em sua forma (terminação). Sendo assim, nossa intenção é investigar se palavras transparentes quanto ao gênero gramatical do português (palavras femininas que terminam em *-a* como *casa*, e masculinas que terminam em *-o* como *carro*) são processadas pelos mesmos mecanismos neurocognitivos que palavras opacas quanto ao gênero (palavras femininas que **não** terminam em *-a* como *noite*, *ordem*, *garagem*, *moto*, e masculinas que não terminam em *-o* como *dia*, *abacaxi*, *tomate*). Para tanto, coletamos neste experimento a atividade cerebral de falantes nativos do português do Brasil para que certos sinais eletrofisiológicos associados a diferentes aspectos do processamento da linguagem possam ser extraídos e analisados.

Nossa hipótese é a de que palavras transparentes e palavras opacas quanto ao gênero gramatical português são processadas por mecanismos neurocognitivos diferentes.

Se você tiver mais perguntas sobre este estudo, não hesite em contactar a pesquisadora principal Natália Resende (natcarol2000@gmail.com). Estamos a sua disposição para maiores esclarecimentos.

Agradecemos sua colaboração!

APÊNDICE I: Protocolo do experimentador

1. Before you start:

- Turn the big red 'stop' button (next to the computers) to turn on the electricity.
- Turn on the computers.
- **Check the audio settings** (see picture). Use -7.5 dB on the amplifier and max. volume in windows on the Presentation computer.
- Turn on the lights inside the experiment room.
- Put ready everything that you need:

Syringes, needles, gel, 6 plastic rings, stickers, alcohol, cotton pads, chin strap, charged batteries, tape-measure, paper with electrode layout...

2. When subject arrives:

- Informed consent
- Measure head size

3. EEG setup

During EEG setup: let subjects read print of instructions

- Put on the cap with all the electrodes, attach the chin strap and make sure the cap is centered and in the correct position.
- Clean the subject's skin where the additional electrodes will be placed with alcohol and a cotton pad.
- Attach the additional electrodes (Ground, References, eye movements)

- Forehead:	GND electrode	
- Left ear:	REF electrode	
- Right ear:	electrode	nr. 32

(from 1-32 set)

- Under the left eye:	electrode	nr. 29
-----------------------	-----------	--------

(from 33-64 set)

- Put gel into all the electrodes, trying to reduce impedance until all the lights are green. **!! Always start with putting**

gel in the ground and reference electrodes.

- Connect a PowerPack (battery) to the two amplifiers.
- Connect the EEG system to the two amplifiers and turn on the two amplifiers (little switches on the back)

4. Starting the EEG experiment:

- Let the subject see his/her EEG signal and show the effect of blinking, moving, etc.

When showing EEG signal. Explain that in my experiment they should NOT blink during sentence reading, but they can blink after the sentence presentation when the symbol ' - - ' appears on the screen.

- **START TO SAVE EEG RECORDING IN VISION RECORDER.**
- **Start practice**
- Start the experiment and use correct subject number. Once scripts started, it takes about 45 mins. for the experiment to end.

5. After the experiment

- Show the subject the way to the basement to clean their hair.
- Give the subject the opportunity to read the debriefing.
- Clean the EEG set:

Scrub all the electrodes with the toothbrush. Pick through the hole in every electrode with help of the other brush. It is very important that you do this, because otherwise the holes will get blocked, and the EEG system cannot be used properly the next time. Clean the cap, make sure there is no more gel in any of the holes. During the cleaning, make sure the splitter boxes do not come into contact with any water!

- Make a copy of the data.
- Write down in the EEG logbook which cap and EEG set you used and if there were any problems/defects/etc.
- Charge the batteries.
- Charge the Powerpack.

- Switch of the lights in the experiment room.
- Turn of all the computers.
- Turn of the electricity

